



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

교육학박사학위논문

중등학교 영어교사의 학생평가 전문성
평가도구 개발 및 타당화 연구

2017년 8월

서울대학교 대학원

교육학과 교육학전공

김 나 영

국 문 초 록

최근 전 세계적으로 미래 사회의 변화에 대응하기 위한 다양한 교육 개혁 정책이 추진되어 왔으며 이러한 교육개혁은 교육평가에 있어서도 많은 변화를 가져왔다. 교육평가의 패러다임이 ‘학습 결과에 대한 평가 (assessment of learning)’에서 ‘학습 향상을 위한 평가(assessment for learning)’ 및 ‘학습 과정으로서의 평가(assessment as learning)’로 변화하면서 교육 현장에서 핵심적인 역할을 수행하는 교사의 새로운 역할과 인식 변화가 요구되고 있다. 우리나라에서도 최근 자유학기제와 성취평가제 및 2015 개정 교육과정의 본격적인 도입에 따라 교수·학습과 평가가 통합된 활동으로서 학생평가의 의미를 이해하고 보다 신뢰도와 타당도가 높은 평가활동을 수행하기 위하여 학교 현장에서 교사의 학생평가 전문성이 강조되고 있다. 특히, 2015 개정 교육과정에서 교수·학습 과정에서의 학생 참여 및 과정 중심 평가가 강조되면서 교사의 학생평가 전문성에 대한 교육적 관심이 증가하였으며 학생평가 활동에 있어서 교사의 실천적 역량이 더욱 주목받고 있다. 이러한 교육적 맥락에서, 실제 학생평가 활동을 반영한 신뢰롭고 타당한 평가도구를 개발하여 교사의 학생평가 전문성 수준을 객관적으로 평가하고 이를 바탕으로 교사의 학생평가 전문성 향상을 위한 지원책을 마련할 필요가 있다.

이를 위해 이 연구에서 중등학교 영어교사의 실천적 학생평가 전문성을 평가할 수 있는 도구를 개발하고 타당화 연구를 수행하고자 한다. 먼저, 국내·외 다양한 문헌 연구 분석 및 전문가 협의회를 거쳐 중등학교 영어교사의 학생평가 전문성 평가척도, 학생평가 포트폴리오의 구성요소 및 전문가 평정을 위한 채점기준을 개발하였다. 다음으로, 단기간의 일회성 평가가 아닌 한 학기 동안의 교사의 실제적인(authentic) 학생평가 활동을 반영하기 위하여 서울, 경기, 인천 지역의 중·고등학교 영어교

사를 대상으로 수집된 ‘학생평가 포트폴리오’를 평가 자료로 활용하였다. 3인의 전문가 집단이 개발한 평가도구를 활용하여 학생평가 포트폴리오에 대한 평가를 실시하였으며, 전문가 평가 실시 후 연구 참여 교사 및 전문가 평가자를 대상으로 평가도구의 적합도에 관한 개방형 설문조사를 실시하였다. 평가 결과를 바탕으로 평가도구의 타당도, 신뢰도 및 적합도에 관한 분석을 실시하였다.

이 연구의 주요 연구 결과를 요약하면 다음과 같다.

첫째, 중등학교 영어교사의 학생평가 전문성 평가척도는 ‘평가계획’, ‘평가도구 개발’, ‘평가 실행’, ‘평가 결과 분석 및 활용’, ‘평가에 대한 태도 및 인식’의 5개의 하위 영역으로 구성되었으며, 각 영역별로 5개의 세부적인 평가 기준을 포함하였다. 하위 영역별로 학생평가 전문성의 평균 점수를 비교해보면, ‘평가도구 개발’ 영역의 평균 점수가 가장 높게 나타났으며, ‘평가 결과 분석 및 활용’과 ‘평가에 대한 태도 및 인식’ 영역의 평균 점수는 상대적으로 낮은 편이었다. 세부 평가 기준별로 살펴보면, ‘평가의 실제성’ 항목의 점수가 가장 높은 반면 ‘평가 참여’와 ‘학부모와의 의사소통’ 항목의 평균 점수가 가장 낮은 것으로 나타났다.

둘째, 영어 교육 전문가 5인을 대상으로 개발한 평가도구의 내용 타당도를 분석한 결과, 25개의 세부 평가 기준은 해당 학생평가 포트폴리오 자료를 활용하여 학생평가 전문성의 다섯 가지 구성 요인을 평가하기에 적절한 내용으로 구성되었음을 확인할 수 있었다.

셋째, 평가도구의 신뢰도를 검증한 결과 비교적 양호한 수준의 신뢰도를 확보한 것으로 나타났다. 고전검사 이론을 활용하여 평가 문항의 내적 신뢰도를 분석한 결과, 하위 영역별로 Cronbach α 계수가 .56~.78이었으며, 전체적으로 .88을 나타냈다. 또한, 전문가 3인의 평정 결과 간의 상관계수는 .84~.98로 나타나 높은 수준의 채점자간 일치도를 보여주었다. 일반화가능도 이론을 활용하여 신뢰도를 분석한 결과, G 계수는 0.8743이며 D 계수는 0.7961으로 각각 0.8과 0.7 이상으로 적정

한 수준의 신뢰도를 확보한 것으로 나타났다.

넷째, 평가도구의 적합도에 관한 개방형 설문조사 분석 결과, 학생평가 포트폴리오를 활용한 평가는 교사들에게 자신의 학생평가 활동 전반에 대한 성찰과 반성의 기회를 제공하며, 학생평가 전문성 향상을 위한 보다 실제적이고 효과적인 방안이 될 수 있다는 점에서 유용한 것으로 나타났다. 반면에 학생평가 포트폴리오의 데이터 수집상의 어려움이나 주어진 포트폴리오 자료만을 활용하여 교사의 학생평가 전문성을 다각적인 측면에서 평가해야 하는 점은 이 평가도구의 한계점으로 지적되었다. 이는 학생평가 포트폴리오의 데이터 수집이나 평정 과정에서 나타나는 일부 한계점을 보완한다면 교사의 학생평가 전문성 평가도구가 교사의 실천적인 학생평가 전문성을 신장시키기 위한 교사 교육 및 교사 연수에서 보다 효과적으로 활용될 수 있음을 시사한다.

이 연구는 연구 대상자 수의 제한으로 인해 연구 결과를 일반화하거나 평가도구의 적합도 측면에서 일부 한계점이 존재하였다. 그럼에도 불구하고, 교사의 학생평가 포트폴리오를 활용한 학생평가 전문성 평가도구는 기존의 평가 관련 교사 연수에서 지적되는 문제점을 극복할 수 있는 매우 효과적인 방안으로 교사의 실천적 평가 전문성 신장을 위해 활용될 수 있는 가능성이 확인되었다.

***주요어:** 학생평가 전문성, 실제적(authentic)학생평가활동, 학생평가 포트폴리오, 평가도구 개발, 타당화 연구

*** 학 번 :** 2013-31074

목 차

I. 서론	1
1. 연구의 필요성 및 목적	1
2. 연구 문제	6
II. 이론적 배경	8
1. 교사의 학생평가 전문성 개관	8
가. 교사의 학생평가 전문성의 개념 및 구성요소	8
1) 학생평가 전문성의 개념	8
2) 학생평가 전문성의 구성요소	9
3) 실천적 역량으로서의 학생평가 전문성	18
4) 영어 교과에서의 학생평가 전문성	21
나. 평가 문해력의 개념 및 필요성	26
1) 평가 문해력의 개념	26
2) 평가 문해력의 필요성	29
다. 교사의 학생평가 전문성 함양 실태	31
2. 교사의 학생평가 전문성 평가도구	33
가. 내용적 측면	34
나. 방법론적 측면	37
다. 선행 연구의 한계점	45

Ⅲ. 연구 방법	47
1. 연구 절차	47
2. 연구 대상	52
가. 연구 참여 교사	52
나. 평가도구의 내용 타당도 검증을 위한 전문가 집단	54
다. 학생평가 포트폴리오 평정을 위한 전문가 평가자	55
3. 자료 분석 방법	56
가. 평가도구의 타당도 검증	56
나. 평가도구의 신뢰도 검증	57
1) 고전검사이론을 활용한 신뢰도 검증	57
2) 일반화가능도이론을 활용한 신뢰도 검증	58
다. 평가도구의 적합도 검증	63
Ⅳ. 연구 결과	64
1. 교사의 학생평가 전문성 평가도구의 개발	64
가. 중등학교 영어교사의 학생평가 전문성 평가척도	65
나. 학생평가 포트폴리오 구성요소	73
다. 전문가 평정을 위한 채점기준	77
2. 교사의 학생평가 전문성 평가도구의 타당화	78
2.1. 기술통계 분석 결과	78
가. 학생평가 전문성의 평가 기준별 기술통계 결과	78
나. 교사 배경변수에 따른 학생평가 전문성의 기술통계 결과	82

1) 성별에 따른 학생평가 전문성 분석 결과	82
2) 교직경력에 따른 학생평가 전문성 분석 결과	84
3) 학력에 따른 학생평가 전문성 분석 결과	85
4) 학교급에 따른 학생평가 전문성 분석 결과	87
2.2. 평가도구의 타당도 분석	89
2.3. 평가도구의 신뢰도 분석	91
가. 고전검사이론 분석 결과	91
나. 일반화가능도이론 분석 결과	93
1) G 연구 분석 결과	93
2) D 연구 분석 결과	95
2.4. 평가도구의 적합도 분석	98
 V. 요약 및 논의	107
1. 요약	107
2. 논의 및 시사점	111
 참고문헌	120
부 록	134
Abstract	181

표 목 차

<표 II-1> 교사의 평가 전문성 기준	10
<표 II-2> 교실 평가 기준	12
<표 II-3> 선행 연구의 학생평가 전문성 기준 및 세부 항목	14
<표 II-4> 교사의 학생평가 전문성 영역 및 기준	16
<표 II-5> 학생평가 전문성의 하위 요소와 점검 내용	17
<표 II-6> 평가 전문성과 평가 문해력의 개념 비교	28
<표 II-7> 초등학교 교사의 학생평가 전문성 척도	40
<표 II-8> 교사의 학생평가 전문성 평가도구의 하위 영역 및 측정 방식	43
<표 III-1> 연구 참여 교사의 표집수	53
<표 III-2> 영어교육 전문가들의 인적 사항	54
<표 III-3> 학생평가 포트폴리오 평가자들의 인적 사항	55
<표 III-4> 연구 모형의 자료 배열	60
<표 IV-1> 중등학교 영어교사의 학생평가 전문성 평가척도	70
<표 IV-2> 학생평가 포트폴리오의 구성요소	76
<표 IV-3> ‘학생 안내’ 항목에 관한 세부 채점기준	77
<표 IV-4> 학생평가 전문성 총점 및 하위 영역별 기술통계 결과	79
<표 IV-5> 학생평가 전문성의 세부 평가 기준별 기술통계 결과	80
<표 IV-6> 성별에 따른 학생평가 전문성의 기술통계 결과	82
<표 IV-7> 교직경력에 따른 학생평가 전문성의 기술통계 결과	84
<표 IV-8> 학력에 따른 학생평가 전문성의 기술통계 결과	86
<표 IV-9> 학교급에 따른 학생평가 전문성의 기술통계 결과	88
<표 IV-10> 문항의 적절성에 관한 내용 타당도 분석 결과	90
<표 IV-11> 평가도구의 문항 내적 신뢰도	92
<표 IV-12> 평정 결과의 채점자간 일치도	92
<표 IV-13> p x i x r 설계에 대한 G 연구 분석 결과	94
<표 IV-14> p X I X R 설계에 대한 D 연구 분석 결과	96

그 립 목 차

[그림 III-1] 연구 절차	47
[그림 III-2] 상대적 오차 및 절대적 오차	59
[그림 III-3] p x i x r 설계의 벤다이어그램	61
[그림 IV-1] 교사의 학생평가 전문성 총점	175
[그림 IV-2] ‘평가계획’ 영역의 점수분포	175
[그림 IV-3] ‘평가도구 개발’ 영역의 점수분포	176
[그림 IV-4] ‘평가 실행’ 영역의 점수분포	176
[그림 IV-5] ‘평가 결과 분석 및 활용’ 영역의 점수분포	177
[그림 IV-6] ‘평가에 대한 태도 및 인식’ 영역의 점수분포	177
[그림 IV-7] 성별에 따른 교사의 학생평가 전문성 평균점수 비교	83
[그림 IV-8] 교직경력에 따른 교사의 학생평가 전문성 평균점수 비교	85
[그림 IV-9] 학력에 따른 교사의 학생평가 전문성 평균점수 비교	87
[그림 IV-10] 학교급에 따른 교사의 학생평가 전문성 평균점수 비교	89
[그림 IV-11] 평가 문항수에 따른 G 계수 및 D 계수의 변화 양상	97

I. 서론

1. 연구의 필요성 및 목적

최근 전 세계적으로 제 4차 산업 혁명으로 인한 미래 사회의 변화에 대응하기 위해 다양한 교육 개혁 정책을 추진해 왔다. 이러한 교육 개혁은 미래사회가 요구하는 핵심역량의 함양을 위한 방향으로 추진되어 왔으며 교육평가에 있어서도 많은 변화를 가져왔다.

교육평가는 ‘교수 프로그램의 교육 효과에 관한 의사결정을 하기 위해서 학습자의 행동변화 및 학습과정에 관한 정보를 수집하고 이용하며 교육적 의사결정을 내리는 데 도움을 주는 과정’ 이라고 정의할 수 있다 (황정규, 1998, p37). 교육평가의 패러다임은 최근 학습 활동과 평가 활동 간의 관계에 따라 ‘학습 결과에 대한 평가(assessment of learning)’ 에서 ‘학습 향상을 위한 평가(assessment for learning)’ 및 ‘학습 과정으로서의 평가(assessment as learning)’ 로 변화하고 있다 (McMillan, 2014). 먼저, ‘학습 결과에 대한 평가’ 는 학생의 학습 수행과정과 그 결과에 대한 의사결정을 내리기 위해 실시되는 평가로서 교사 중심의 총합평가(summative assessment)의 성격이 강하다. 반면에, ‘학습 향상을 위한 평가’ 의 경우, 교사가 학생평가 정보를 통해 학생의 학습과정을 지원하고 돕는 형성평가(formative assessment)의 특성을 지닌다. 나아가, ‘학습 과정으로서의 평가’ 는 학습자가 자기 평가와 동료 평가 방식을 통해 스스로를 평가하고 평가 정보를 활용하여 ‘자기 주도적인 학습을 이끌 수 있도록(self-regulation)’ 학습과 평가가 통합된 방식을 의미한다(김성숙 외, 2015; 성태제 외, 2015). 평가활동

을 수업활동과 분리된 것으로 인식하는 지필평가 중심의 전통적 평가 방식에서 벗어나 교수·학습상황에서 학생평가가 자연스럽게 수업과의 연계성을 갖게 되면서 단위학교에서 교육의 질 향상을 위해 학습 지향적 평가로서의 형성평가, 학생들의 문제해결능력을 실제적, 다면적 방법으로 평가하는 수행평가, 참평가, 포트폴리오, 저널쓰기, 기술 공학 기반 문항 등 학생평가 활동에서 대안적 평가 방식의 중요성이 더욱 강조되고 있다(김성숙 외, 2015; Brookhart, 2007; McMillan, 2014; Stiggins, 2005). 이와 같은 교육평가 패러다임의 전환은 교육 현장에서 핵심적인 역할을 수행하는 교사의 새로운 역할과 인식 변화에도 시사하는 바가 크다고 할 수 있다.

이러한 교육평가의 흐름에 발맞추어 우리나라에서도 최근 자유학기제와 성취평가제 및 2015 개정 교육과정의 본격적인 도입에 따라 교사의 학생평가 활동에 자율성이 확대되고 교사의 학생평가 전문성 신장에 대한 교육적 관심이 높아졌다. 먼저, 자유학기제에서는 학생평가의 기본 방향으로 ‘평가 방법의 다양화’, ‘평가 목적의 정상화’, ‘평가 영역의 균형화’, ‘평가 주체의 다변화’ 등 크게 네 가지 요소를 강조하고 있다(김진규, 2013). 성취평가제에서는 성취기준에 따른 교수·학습 및 평가가 강조되고 개별 학생들의 능력에 대한 정확한 진단과 향상을 목표로 함으로써 학생평가에서 교육과정, 교수·학습, 평가의 연계가 이루어지고 있다(김경희 외, 2012; 박지현 외, 2017). 다음으로, 2015 개정 교육과정에서는 교수·학습 과정에서의 학생 참여 및 과정 중심 평가가 강조되면서 교실 내 협력학습, 프로젝트 학습 등 다양한 형태의 수행평가와 형성평가의 시행이 확대되고(교육부, 2015), 평가 과정에서의 학생 참여가 더욱 활발해지고 있다. 또한, 교사가 학생들에게 학습 결과에 대한 정보 제공에서 나아가 교수·학습 과정에 관한 개별적인 피드백 제

공이 확대되는 등 학생평가 활동에 있어서 교사의 실천적 역량이 더욱 강조되고 있다.

따라서, 교사의 학생평가 역량의 강화는 교실 수업 개선 및 교육의 질 제고와 직결된다는 점에서 신뢰롭고 타당한 평가도구를 개발하여 교사의 학생평가 전문성 수준을 객관적으로 평가하고 이를 바탕으로 평가 관련 교사 연수를 비롯한 교사의 평가 전문성 향상을 위한 지원책을 마련한 필요성이 있다.

대부분의 선행연구에서는 한국교육과정평가원에서 수행한 송미영, 김경희(2007)의 ‘교사의 학생평가 전문성 진단도구’를 활용한 자기 평가 방법을 통해 교사의 학생평가 전문성을 평가하고자 하였다. 하지만, ‘교사의 학생평가 전문성 진단도구’는 10년 전에 개발되어 최근의 교육과정이나 교육평가의 맥락을 반영하지 못하였으며, 다음과 같은 몇 가지 한계점을 지니고 있다.

첫째, 주로 평가 지식 및 기능 등의 이론적 측면에 초점을 두어 교육 현장에서 실제 학생평가 활동을 반영한 교사의 실천적 평가 전문성을 평가하는 데에는 어려움이 있었다. 둘째, ‘평가의 윤리성’ 항목을 평가 구성요소에 포함하고는 있으나 ‘교사의 평가 전문성 개발 노력’이나 ‘동료교사와 협업능력’, ‘평가의 환류적 기능’ 등을 포함하여 평가의 정의적 영역에 관한 보다 포괄적인 고려가 부족하였다. 셋째, 자기 보고식 설문조사를 바탕으로 한 자기 평가 방법을 활용하여 교사의 학생평가 전문성에 관한 보다 객관적인 측정이 어렵다는 문제점이 지적되었다.

따라서, 이 연구에서는 이와 같은 한계점을 극복하기 위해 교사의 학생평가 전문성을 실제적으로 평가할 수 있는 새로운 평가도구를 개발하고, 학생평가 포트폴리오를 활용한 전문가 평가방법을 실시하여 개발된 평가도구의 타당도, 신뢰도 및 적합도를 확인하고자 한다. 평가도구 개

받을 위한 구체적인 방향을 제시하면 다음과 같다.

첫째, 교사의 학생평가 전문성의 실천적 역량을 평가하기 위하여 평가 계획에서부터 평가도구를 개발하고 평가를 실행한 후 평가 결과에 관한 학생 피드백의 제공 등 교사의 실제적인 학생평가 활동의 모습인 ‘학생평가 포트폴리오’를 평가 자료로 활용하고자 한다. 이는 2015 개정 교육과정에서 추구하는 과정 중심 평가에 유리할 뿐만 아니라 단기간의 일회성 평가가 아닌 한 학기 동안의 교사의 실제적인(authentic) 학생평가 전문성을 평가할 수 있다는 점에서 의의를 지닌다.

둘째, 학생평가 전문성의 인지적인 요소뿐만 아니라 정의적 영역을 다각도로 고려하여 교사의 학생평가 전문성을 보다 포괄적으로 평가할 수 있는 도구를 개발하고자 한다.

셋째, ‘자기 평가 방법’을 활용할 경우 스스로를 관대하게 평가하는 경향이 있음을 볼 때(어윤경, 2006; 황은희, 백순근, 2008), ‘전문가 평가 방법’을 활용하여 보다 타당하고 객관적인 평가가 이루어질 수 있도록 한다.

넷째, 교사의 잠재적 실천 역량은 상황 또는 영역 특수성(domain specificity)을 지님을 볼 때(백순근, 임현수, 2006), 교사의 학생평가 전문성의 과목 특수적, 맥락 의존적 특성을 고려하여(Xu & Brown, 2016) 평가도구 개발 시 교과목의 세부적인 특성과 평가 대상을 보다 구체화하고자 한다.

다섯째, ‘교사의 학생평가 전문성 진단도구’는 주로 교사의 학생평가 전문성의 현재 역량을 기준 점수와 비교하여 진단하는 데 그 목적이 있었다. 새로운 평가도구는 단순히 교사의 현재 역량을 진단하는데 그치지 않고 교사의 학생평가 전문성 신장에 효과적인 도구로 개발할 필요가 있다. 학생평가 포트폴리오를 활용한 평가의 경우 교사의 개별적인 학생

평가 전문성의 구체적인 특징, 장점 및 개선할 점 등을 일목요연하게 파악할 수 있 수 있으며 자기반성과 성찰의 기회를 제공한다는 점에서 교사의 학생평가 전문성 향상에 실질적인 도움을 줄 것으로 기대된다.

이와 같은 기본 방향에 따라 이 연구에서는 중등학교 영어교사를 대상으로 수집한 학생평가 포트폴리오를 활용하여 교사의 실제적인 학생평가 전문성을 평가할 수 있는 도구를 개발하고 타당화하고자 한다. 이 연구는 교사가 스스로의 학생평가 전문성을 보다 객관적으로 파악하는데 도움을 줄 뿐만 아니라 학생평가 전문성 함양을 위한 교사 교육 및 교사 연수에 중요한 기초자료를 제공할 것이다.

2. 연구 문제

이 연구는 선행연구에서 발견되는 평가도구의 타당도와 신뢰도의 한계점을 극복하기 위해 교사의 실천적 학생평가 전문성을 측정할 수 있는 평가도구를 개발하고 개발된 평가도구를 활용하여 중등학교 영어교사의 학생평가 포트폴리오에 대한 전문가 평가를 실시하고자 한다. 전문가 평가 결과 분석을 통해 평가도구의 타당도와 신뢰도를 검증하고 연구 참여 교사 및 전문가 평정자를 대상으로 개방형 설문조사를 실시하여 평가도구의 적합도를 검증할 것이다.

교사의 실천적 학생평가 전문성을 평가할 수 있는 도구를 개발하고 타당화하기 위한 이 연구의 주요 내용을 제시하면 다음과 같다. 첫째, 평가도구 개발에 앞서 교사의 학생평가 전문성의 평가 기준에 관해 살펴볼 필요가 있다. 이를 위해 관련 선행연구 분석을 통해 학생평가 전문성의 구성요소를 선정하고 세부 평가 기준을 개발할 것이다.

둘째, 실제 중등학교 영어교사들을 대상으로 수집한 데이터를 바탕으로 개발한 평가도구의 타당도를 분석하고자 한다. 이를 위해 영어 교육 전문가 집단을 대상으로 중등학교 영어교사의 학생평가 전문성 평가척도의 각 평가 문항이 학생평가 전문성의 하위 구성요소를 제대로 평가하고 있는지에 관한 내용타당도 분석을 실시할 것이다.

셋째, 개발한 평가도구의 신뢰도를 보다 정확하게 검증하기 위해 학생평가 포트폴리오에 관한 3인의 전문가 평가자의 평정 결과를 바탕으로 고전검사이론과 일반화가능도이론을 각각 활용하여 분석을 실시하고자 한다.

넷째, 교사의 학생평가 포트폴리오를 활용한 학생평가 전문성 평가도

구가 평가의 목적이나 취지에 적합한지를 검증하고자 한다. 이를 위해 연구 참여 교사 및 전문가 평정자를 대상으로 평가도구의 적합도에 관한 개방형 설문조사를 실시하고 응답 내용을 구체적으로 분석할 것이다.

이를 위해 설정한 주요 연구문제를 요약하면 다음과 같다.

1. 중등학교 영어교사의 학생평가 전문성의 평가 기준은 무엇인가?
2. 중등학교 영어교사의 학생평가 전문성 평가도구의 내용 타당도는 적합한가?
3. 중등학교 영어교사의 학생평가 전문성 평가도구의 신뢰도는 적합한가?
4. 중등학교 영어교사의 학생평가 전문성 평가도구는 평가의 목적과 취지에 적합한가?

Ⅱ. 이론적 배경

1. 교사의 학생평가 전문성 개관

가. 교사의 학생평가 전문성의 개념 및 구성요소

교사의 학생평가 전문성 평가도구를 개발하기에 앞서 선행연구를 바탕으로 학생평가 전문성의 개념 및 구성요소에 관해 고찰해 볼 필요가 있다.

1) 학생평가 전문성(Teacher's Competency in Student Assessment)의 개념

교육의 내실화 및 학교 교육의 질적 제고에 대한 요구가 증가하면서 교사 전문성은 학교 교육의 효과성에 영향을 미치는 가장 중요한 변수로 강조되고 있다(손승남, 2005; 이종일, 2003; 이혜정, 2010). 아울러 학습자의 목표 설정 및 발달적 측면에서도 실제 교육 현장에서 학생들과 직접적으로 상호작용하는 교사의 역할 및 전문성이 주목받고 있다(신중호 외, 2012). 교사 전문성과 관련된 주요 선행 연구를 살펴보면, 교사 전문성의 개념모형을 설정한 연구(김옥예, 2006; 소경희, 2003)와 교사의 수업 능력 개발 요구에 따라 교사의 수업 전문성을 분석한 연구(김정환, 이계연, 2005; 손승남, 2005)가 대부분이었다.

미국의 NCLB (No Child Left Behind)정책 실시 이후, 교육의 책무성에 대한 관심이 증가되면서 학생들의 학업 성취를 향상시키기 위한 방안으로 교사 전문성 중 특히 교사의 평가 전문성 함양을 강조하고 있다(DeLuca & Klinger, 2010). 교사의 평가 활동이 업무 시간 또는 수업

과정의 절반 이상을 차지하며(김신영, 2002 ; Plake, 1993; Mertler, 2005b), 교사의 학생평가 전문성이 교사의 교수 활동 개선과 학생의 학업 성취에 영향을 준다는 연구 결과를 바탕으로 할 때, 교사의 학생평가 전문성에 대한 관심이 더욱 부각되고 있다(이인제 외, 2004; 전병만 외, 2005; 지은림 외, 2011; Shohamy, 2001; Stiggins, 1999). 즉, 교사의 평가 전문성은 ‘교수효과 및 학습효과의 증대’, ‘학교교육의 효과 증대(교육목표 달성도 증대)’, ‘학교 교육 기능의 정상화’, ‘교육의 질적 향상 노력’ 과 관련된 매우 핵심적인 교육활동으로 교육평가의 성과가 교사의 평가 전문성에 달려있다고 해도 과언이 아니다(배호순, 2008).

이러한 중요성에도 불구하고, 교사의 학생평가 전문성에 관한 개념이나 구성요소에 관한 합의된 연구는 활발히 이루어지지 않았다(김선희, 2006; 김진완 외, 2004). 이 연구에서는 교사의 학생평가 전문성을 ‘학생의 학습과 성취에 관한 다양한 평가 정보를 수집하고 해석하여 활용할 수 있는 능력’ 으로 정의하였으며 평가 활동을 수행하기 위해 교사에게 요구되는 핵심 역량을 의미하는 평가 소양(assessment literacy for teacher)으로 해석하여 기술하였다(김경희 외, 2006; 김수동 외, 2005a; Brown, 2004; DeLuca & Klinger, 2010).

2) 학생평가 전문성의 구성요소

먼저, 교사의 학생평가 전문성의 능력 요소 및 기준과 관련된 국외 연구를 살펴보면 다음과 같다. 미국교사연합회(American Federation of Teacher: AFT), 교육측정위원회(National Council on Measurement

in Education: NCME), 국립교육연합회(National Education Association: NEA)는 1990년도에 학생평가 활동을 위해 교사에게 필요한 7개의 능력 요인으로 구성된 ‘교사의 학생평가 전문성 기준(Standards for Teacher Competence in Educational Assessment of Students)’을 제시하였다. 이 기준은 ① ‘평가 방법의 선정’ ② ‘평가 방법의 개발’ ③ ‘평가의 실시, 채점 및 결과 해석’ ④ ‘평가 결과의 활용’ ⑤ ‘성적 부여’ ⑥ ‘결과에 관한 의사소통’ ⑦ ‘윤리적 문제의 인식’으로 구성된 것으로 주로 교사들의 검사(testing)와 관련된 원리와 지식을 중점적으로 다루고 있어 포괄적인 측면에서 평가 전문성을 다루기에는 한계점이 존재하였다(김신영, 2002; 박가나, 2012).

Brookhart(2011)는 AFT et al.(1990)의 ‘교사의 학생평가 전문성 기준’이 형성평가 개념을 포함하지 않으며 기준 중심 교육개혁(standards-based reform)과 교육적 책무성 개념을 고려하고 있지 않다는 점을 지적하였다. 이에 Brookhart(2011, p7)는 최근의 교육적 이슈를 반영하여 학생평가의 다각적 측면을 고려한 새로운 평가 전문성 기준을 <표 II-1>과 같이 11가지로 제시하였다.

<표 II-1> 교사의 평가 전문성 기준 (Brookhart, 2011)

영역	기준	내용
평가 준비	내용의 이해	교사는 자신이 가르치는 내용 영역의 학습에 대해 이해해야 한다.
	교육과정의 이해	교사는 교육과정에 근거하여 분명한 학습 의도를 달성가능하며, 평가 가능한 방식으로 나타낼 수 있어야 한다.
	성취수준 설정	교사는 학습의도에 따른 성취수준이 무엇인지 학생들과 의사소통할 수 있는 전략을 지니고 있어야 한다.
	효과적인 평가 방법	교사는 다양한 평가 방법의 목적 및 용도를 이해하고, 또한 이들을 숙련되게 활용할 수 있어야 한다.

평가 실행	평가문항 개발	교사는 학생들에게 필요한 특정한 지식과 사고 기술을 확 인할 수 있도록 수업 중의 질문이나 평가 문항 및 수행평 가 과제를 분석할 수 있는 능력을 지니고 있어야 한다.
	피드백 제공	교사는 학생의 성취도에 대해 효과적이고 유용한 피드백 을 제공 할 수 있는 능력을 갖추어야 한다.
	채점기준의 개발	교사는 학생의 학습, 성장 및 발전을 개선하기 위한 평가 기준을 구성할 수 있어야 한다.
평가 결과 분석	평가 결과의 해석	교사는 외부평가를 시행할 수 있고, 학생, 교실, 학교 및 지역 수준에서 해당 평가결과를 해석할 수 있어야 한다.
	평가 결과에 대한 의사소통	교사는 관련 교육적 집단(학생, 학부모, 학급, 학교 공동 체)에게 평가 결과를 해석하고, 평가 결과에 기반한 교육 적 의사결정에 대한 설명을 분명하게 할 수 있어야 한다.
	평가 결과의 활용	교사는 학생들의 공정한 교육적 의사결정을 위하여 평가 정보를 사용할 수 있도록 도울 수 있어야 한다.
평가의 윤리성	윤리적 책임	교사는 평가 업무와 관련된 법적, 윤리적인 책임을 이해 하고 이행해야 한다.

Brookhart(2011)의 기준은 교사가 단순히 평가도구를 개발하고 평가를 실행함에 있어서 필요한 지식과 기술에서 나아가 평가 준비에서부터 교육적 의사결정을 위해 평가 결과를 활용하며 교육 당사자들과 의사소통할 수 있는 능력을 포함하는 등 보다 포괄적인 학생평가 전문성을 다루고 있다는 점에서 기존의 평가 전문성 기준들과 구별된다.

다음으로, 교육평가 표준 공동 위원회(The Joint Committee on Standards for Educational Evaluation, JCSEE, 2015)의 교실평가 기준(Classroom Assessment Standards)은 ‘평가 기초 기준’, ‘평가 실행 기준’ 및 ‘평가의 질에 관한 기준’의 총 3개 영역에서 16개의 세부 기준을 제시하였다. 구체적인 교실 평가 기준을 살펴보면 <표 II-2>와 같다.

<표 II-2> 교실 평가 기준 (JCSEE, 2015)

영역	세부 기준	내용
평가 기초 기준 (Foundation Standards)	평가목적 (Assessment Purpose)	교실 평가 활동은 교수·학습을 지원할 수 있는 명확한 평가 목적을 지녀야 한다.
	학습 기대/목표 (Learning Expectations)	학습목표는 교실 평가 활동이 개별 학생들을 위한 적절한 수업 및 학습 기회와 부합되도록 수업, 학습 및 평가의 기초를 형성해야 한다.
	평가 계획 (Assessment Design)	활용된 교실 평가의 유형과 방법은 학생들이 스스로의 학습을 명확하게 보여줄 수 (demonstrate) 있어야 한다.
	학생의 평가 참여 (Student Engagement in Assessment)	학생평가는 학습의 향상을 위한 방향이어야 하며 학생들은 평가 과정에 유의미하게 참여해야 한다.
	평가 준비 (Assessment Preparation)	학습 자료, 시간, 학습 기회의 측면에서 교사와 학생의 적절한 준비가 요구된다.
	학생 및 학부모와의 의사소통 (Informed Students and Parents/Guardians)	평가의 목적과 활용에 관해 학생 및 학부모와 의사소통해야한다.
평가실행기준 (Standards of Use)	학생 성취에 관한 분석 (Analysis of Student Performance)	학생들의 학습에 관한 분석 방법은 평가 목적 및 실행에 적합해야 한다.
	효과적인피드백 (Effective Feedback)	교실 평가 활동은 학생들의 학습을 향상시키기 위해 적절한 시기에 유용한 피드백을 제공해야 한다.
	수업개선활동 (Instructional Follow-Up)	학생들의 지속적인 학습개선을 위해 학습 성취에 관한 분석을 향후 수업 계획에 반영해야 한다.
	성적 부여 및 총평 (Grade and Summary Comments)	성적부여 시 학습 기대(목표)에 대한 학생들의 성취 정도를 반영해야 한다.
	보고 (Reporting)	평가 보고는 충분한 증거(evidence)에 바탕을 두어야 하며 학생들의 학습 전반에 관해 명확하고, 시기적절하며, 정확하고, 유용한 방법으로 이루어져야한다.

평가의 질에 관한 기준 (The standards of quality)	문화적, 언어적 다양성 (Cultural and Linguistic Diversity)	교실 평가 활동은 학생들의 문화적, 언어적 다양성을 존중해야 한다.
	예외성과 특수 교육 (Exceptionality and Special Education)	교실 평가 활동은 모든 학생들의 구체적인 교육적 요구를 충족시키기 위해 적절하게 다양화, 차별화(differentiated)되어야 한다.
	공정한 평가 (Unbiased and Fair Assessment)	교실 평가 활동과 차후 의사결정은 원래의 평가 목적과 관련 없는 요인에 의해 영향을 받아서 안된다.
	신뢰도와 타당도 (Reliability and Validity)	교실 평가 활동은 모든 학생들의 지식과 기 술에 대해 올바른 해석과 의사결정을 위해 일관성 있고 믿을만하며 적절한 정보를 제 공해야 한다.
	반성 (Reflection)	교실 평가 활동은 전반적인 평가의 질 향상 을 위해 점검되고 개선되어야 한다.

JCSEE(2015)의 기준은 교사의 교실평가 활동을 위한 보다 구체적인 지침을 제공하였다는 점에서 의의를 지닌다.

DeLuca 외(2016)의 연구는 앞서 살펴본 AFT et al.(1990), Brookhart(2011)와 JCSEE(2015)의 학생평가 전문성 기준을 포함하여 영어권 국가 5개국(호주, 캐나다, 뉴질랜드, 영국, 미국)과 유럽 국가에서 개발된 총 15개의 기준을 비교·분석하여, 교사의 학생평가 전문성을 공통적인 8개의 구성요소로 요약하였다(<표 II-3> 참조).

<표 II-3> 선행 연구의 학생평가 전문성 기준 및 세부 항목
(DeLuca et al., 2016, p.255)

기준 영역	내용	세부 항목
평가 목적	명확한 수업 목표에 따라 적합한 평가 형태를 선택할 수 있음	<ul style="list-style-type: none"> • 다양한 형태의 평가 • 형성 평가 vs 총괄평가 • 교실기반평가 vs 표준화된 평가
평가 과정	평가를 구성, 실행, 채점하고 수업을 위한 의사 결정에 도움이 되는 방향으로 평가 결과를 해석할 수 있음	<ul style="list-style-type: none"> • 평가도구의 개발 • 평가 실행 • 평가 정보 수집 • 채점 • 평가 결과의 해석 • 의사결정을 위한 평가 결과의 활용 • 타당한 성적 부여 절차의 개발 • 평가 과정의 모니터링 및 수정
평가 결과에 관한 의사소통	평가 목적, 과정, 결과에 관해 의사소통할 수 있음	<ul style="list-style-type: none"> • 평가 목적 및 과정에 관한 의사소통 • 성적 부여 과정에 관한 설명 • 학부모 및 교육 이해자와의 평가 결과에 관한 의사소통
평가 공정성	학습자들의 다양성을 고려하여 특수 학습자들이 공정한 평가 조건 내에서 평가를 실시하도록 함	<ul style="list-style-type: none"> • 공정한 평가 활동 이행 • 학생들의 성취에 관한 공정한 평가결과의 제공 • 예외적인 학습자에 대한 배려 • 학습자들의 다양성 인지
평가 윤리	평가에 관한 정확한 정보를 제공하며 평가받는 학생들의 권리와 사생활을 보호함	<ul style="list-style-type: none"> • 정확하고 균형적인 정보 제공 • 학습자 권리 및 사생활 보호 • 편의(biases)의 최소화 • 규정에 따름
측정 이론	평가의 심리 측정적인 측면 (psychometric properties)을 이해할 수 있음	<ul style="list-style-type: none"> • 신뢰도 • 타당도 • 대규모 평가 • 규준(norms)과 기준(standards)
학습을 위한 평가	교사의 교수·학습 활동 중 형성 평가를 사용할 수 있음	<ul style="list-style-type: none"> • 형성평가 • 진단평가 • 학습을 위한 평가(assessment for learning) • 자기평가 및 동료평가 • 학생들에게 피드백 제공 • 평가 활동에의 학생 참여
교사 교육 및 지원	교사의 평가 전문성을 개발할 수 있음	<ul style="list-style-type: none"> • 대규모 평가 이용자 교육 • 교사의 평가 전문성 개발 기회 제공 • 교사의 학생평가 활동 지원

다음으로 학생평가 전문성에 관한 국내 선행연구를 살펴보면 다음과 같다. 김신영(2002)의 연구에서는 교사의 평가 전문성의 구성요소를 교실 내 평가의 복잡성을 고려해 ① ‘평가의 교실 내 활용 능력’ ② ‘평가하고자 하는 특성에 대한 이해 능력’ ③ ‘건전한 평가의 질에 대한 이해 능력’ ④ ‘다양한 평가도구에 대한 이해 능력’ ⑤ ‘교실 내 평가 상호관계적인 측면에 대한 이해 능력’ ⑥ ‘결과의 활용 방법에 대한 이해 능력’의 6가지로 제시하였다.

‘교사의 학생평가 전문성 신장 모형과 기준’ (이인제 외, 2004) 연구에서는 교사의 학생평가 전문성 기준을 크게 ‘내용 기준(content standards)’과 ‘수행 기준(performance standards)’으로 범주화하였다. 구체적으로 내용 기준은 ① ‘교육 과정에 대한 지식’ ② ‘교수·학습에 대한 지식’ ③ ‘교육평가에 관한 지식’ ④ ‘교과 내용에 대한 지식’으로 구성되며, 수행 기준에는 ⑤ ‘평가 계획 및 준비 능력’ ⑥ ‘평가 시행능력’ ⑦ ‘평가 결과 분석 및 해석 능력’ ⑧ ‘평가 결과의 보고 및 활용 능력’ ⑨ ‘윤리적 문제 인식과 대처 능력’ 등이 포함되었다.

이인제 외(2004)의 연구 결과를 바탕으로 한국교육과정평가원에서 수행한 김경희 외(2006)의 연구에서는 미국 AFT et al.(1990)에서 제시한 ‘학생평가 전문성 기준 영역’과 제7차 교육과정 해설서의 평가 지침을 반영하여 ‘교사의 학생평가 전문성 기준’을 개발하였다. 구체적으로 학생평가 전문성의 구성요소를 살펴보면, ① ‘평가 방법의 선정 능력’ ② ‘평가도구의 개발 능력’ ③ ‘평가의 실시·채점 및 성적 부여 능력’ ④ ‘평가 결과의 분석·해석·활용·의사소통 능력’ ⑤ ‘평가의 윤리성 인식 능력’의 5개의 기준 영역으로 이루어졌다. 구체적인 학생평가 전문성의 기준 및 평가 지식의 내용을 제시하면 <표 II-4>와 같다.

<표 II-4> 교사의 학생평가 전문성 영역 및 기준 (김경희 외, 2006)

기준영역	학생평가 전문성 기준	평가지식 및 기술의 수
1. 평가방법의 선정	1-1.교사는 학습목표와 평가의 목적을 확인하고 명료화해야 한다.	12
	1-2.교사는 학습목표 및 평가의 목적에 적합한 평가 방법을 선정해야 한다.	
2. 평가도구의 개발	2-1.교사는 평가의 목적과 내용에 적합한 평가도구를 개발하거나 선택하여 사용해야 한다.	8
	2-2.교사는 평가도구의 질을 스스로 점검하고 개선해야 한다.	
3. 평가 실시 · 채점 및 성적 부여	3-1.교사는 평가계획에 부합하도록 평가를 실시해야 한다.	9
	3-2.교사는 정확하게 채점하고 평가의 목적에 부합하도록 성적을 부여해야 한다.	
4. 평가결과의 분석 · 해석 · 활용 · 의사소통	4-1.교사는 평가결과를 정확하고 타당하게 분석 및 해석해야 한다.	19
	4-2.교사는 평가결과를 수업 및 학생에 대한 교육적 의사 결정에 활용해야 한다.	
	4-3.교사는 학생, 학부모, 교육 관련자와 평가 결과에 대해 정확하게 의사소통해야 한다.	
5. 평가의 윤리성	5-1.교사는 학생의 인격을 존중하며 모든 평가활동 시 윤리적 · 법적 책임을 준수해야 한다.	7
	5-2.교사는 학생의 특성 및 배경을 고려하여 공정하게 평가를 실시하고, 평가의 적절성을 판단해야 한다.	
합계	11	55

‘교사의 학생평가 전문성 기준’ (김경희 외, 2006)은 교사가 학생 평가를 계획, 실행하기 위해서 필수적인 지식, 기능, 태도를 구체적으로 제시하고 있다는 점에서 예비 교사 및 현직 교사의 평가 전문성 신장을 위한 기초 자료로서의 의의를 지닌다고 할 수 있다.

김신영(2014)은 메타평가를 위한 교육 본위성 기준에 대한 김성훈

외(2009)의 연구와 김신영(2010)의 연구 결과를 바탕으로 학생평가의 하위 요소와 점검 내용을 다음과 같이 제시하였다(<표 II-5> 참조).

<표 II-5> 학생평가 전문성의 하위 요소와 점검 내용 (김신영, 2014, P145)

하위 요소	점검 내용
학습태도	1) 학생이 교사에 대한 신뢰를 기초로 수업에 집중하고 있는지를 점검하였는가? 2) 학생이 현재 수준에 안주하지 않고 최선을 다하려는 노력을 하고 있는지를 확인하였는가? 3) 학생의 학습활동에 대한 만족도를 확인하고 만족도 향상을 위한 방안을 모색하였는가?
학습 성과의 본질성	1) 학생의 학업성취수준과 향상된 정도를 평가하였는가? 2) 학생 스스로 자신의 학업성취수준과 변화정도를 확인해 보게 하였는가? 3) 학생이 학습과 교사에 대한 긍정적인 태도(자신감과 흥미, 교사에 대한 신뢰 및 존경심 등)를 지니게 되었는지를 확인하였는가?
학습방법 및 환경의 적절성	1) 수업내용과 학습자의 특성·수준에 적합한 학습 자료와 학습 방법을 활용하고 있는지 점검하였는가? 2) 단계적·순차적 접근을 하고 있으며 학습 경험의 배열이 적합한지 확인하였는가? 3) 학생이 수업에 적극적으로 참여하고 교사 또는 동료 학생들과 학습과정과 결과에 대해 역동적으로 상호작용하고 있는지를 확인하였는가? 4) 학생의 학습속도와 학습스타일, 그리고 학생수준에 맞는 학습활동을 전개해가고 있는지 확인하였는가? 5) 학생들이 자발적으로 학습에 참여하고, 학습과정에 대한 충실한 모니터링이 이루어지고 있는지 확인 하였는가? 6) 학습활동의 활성화를 촉진하는 수업환경이나 수업지원 체계가 효율적으로 작동하고 있는지를 확인하였는가? 7) 학업성취(achievement), 향상(progress), 노력(effort) 등을 격려해주고 지원하는 수업분위기 및 교육풍토가 조성되고 있는지를 확인하였는가?

학생평가 전문성의 하위 요소 및 점검 내용을 살펴보는 것은 교사의 학생평가 전문성에 꼭 포함되어야 할 구성요소 및 세부 평가 기준의 개발에 있어서 도움을 줄 것이다.

3) 실천적 역량으로서의 학생평가 전문성

다음으로, 학생평가 전문성에 관한 다양한 국내 선행 연구를 분석대상을 기준으로 분류하면, 예비 교사들의 평가 전문성 신장을 위한 연구(김선희, 2006; 김선희, 2012; 김신영, 2007; 김신영, 송미영, 2008; 박정, 2013)와 현직 교사들을 대상으로 한 연구(권순달, 2010; 김동영, 2006; 박가나, 2012; 서태열 외, 2005)로 나눌 수 있다.

먼저, 예비 교사를 대상으로 한 연구를 살펴보면 다음과 같다. 김선희(2006, 2012), 김신영(2007), 김신영과 송미영(2008), 박정(2013)의 연구는 ‘교사의 학생평가 전문성 기준’ (김경희 외, 2006)의 다섯 가지 구성요소를 토대로 예비 교사의 학생평가 전문성을 분석하였다.

김선희(2006)는 중등학교 수학교사 11명을 대상으로 반 구조화된 면담 방법을 활용하여 평가 활동에 있어서 수학 교사들이 겪는 현실적인 어려움과 평가 전문성 기준에 대한 이해 정도를 바탕으로 효과적인 예비 교사 교육을 위한 시사점을 도출하였다. 김신영(2007)은 예비 교사들의 평가 전문성이 현직 교사들에 비해 모든 영역에서 낮게 나타난 분석 결과에 주목하였다. 교육 현장에서의 실천적 경험을 강조하면서 예비 교사들의 평가 전문성을 신장시키기 위해 평가 관련 지식, 기능 및 태도를 배우고 내면화 할 수 있는 체계적인 교육 기회를 제공해야 한다고 주장하였다. 이러한 맥락에서 교사 양성 과정의 현장 적합성을 높이기 위한

방안으로 김신영과 송미영(2008)은 예비 교사 교육에서 교육 실습을 강조하였다. 그러나 박정(2013)의 연구에서는 교육실습이 예비교사들의 평가 전문성에 미치는 효과가 없다고 나타나 상반된 연구 결과를 보여주었다. 김선희(2012)는 교육 실습과 더불어 학습 일지(교사 포트폴리오) 작성을 통한 반성 중심의 교사 교육이 수학 예비 교사들의 평가 전문성 신장에 도움을 준다는 분석 결과를 바탕으로 현직 교사들을 위한 평가 연수에도 이를 활용할 것을 제안하였다.

다음으로, 현직 교사를 대상으로 한 연구의 경우, 교사의 학생평가 전문성 기준이 실제 학교 현장에서 어떻게 구현되는지를 교과목 특성을 고려하여 분석하였다. 구체적으로 살펴보면, ‘교사의 학생평가 전문성 기준(김경희 외, 2006)’의 다섯 가지 능력 요소를 바탕으로 사회과 교사들의 학생평가 전문성을 분석한 박가나(2012)의 연구와 과학과 교사들의 평가 전문성을 분석한 김동영(2006)의 연구가 있다. 중등학교의 비교과 교사를 대상으로 한 권순달(2010)의 연구에서는 ① ‘평가 이론’ ② ‘평가 방법’ ③ ‘평가 수행’ ④ ‘평가 해석’의 4개의 하위요인을 토대로 보건 교사와 일반 교사들의 학생평가 전문성을 비교·분석하였다.

하지만, 서태열 외(2005)는 ‘교사의 학생평가 전문성 기준’이 교육 평가에서 요구되는 당위성만을 포괄적으로 나열한 것으로 각 기준 간의 체계적인 위계성이 부족하여 개념 모형으로서의 역할을 할 뿐 실천을 위한 절차 모형으로서 활용하는데 한계점이 있다고 지적하였다. 서태열 외(2005)의 연구에서는 지리 평가가 구현되는 절차적 계열성을 중심으로 학생평가 전문성의 기준을 ① ‘교육평가 원리 일반에 관한 이해’ ② ‘교과 내용 영역에 관한 체계적 이해’ ③ ‘내용 영역에 적합한 다양한 평가 도구 선정 및 개발’ ④ ‘교실 현장의 교과 학습 평가 실무 능력’ ⑤ ‘평

가 정보의 활용 능력'의 5가지로 제시하였다.

최근에는 교사 전문성에 관한 활발한 논의가 '교과 교육을 성공적으로 수행하기 위해 필요한 교사의 지식, 기능 및 태도의 총체'를 의미하는 교수 역량(teaching competence)에 대한 관심으로 이어졌는데, 이러한 관점에서 교사의 실천적 수행 측면에 초점을 둔 '실천적 교수 역량'의 중요성이 또한 부각되었다(백순근 외, 2007; 황은희, 백순근, 2008). '실천적 교수 역량'의 측면에서 교사의 학생평가 전문성을 살펴보면, 교사는 단순히 평가도구를 개발하고 평가 결과를 해석하는데 그치지 않고, 학생과의 상호 작용, 수업 활동 계획 및 조직, 성의와 열의 등 학습의 전 과정에서 통합된 활동에 있어서 자질과 능력이 요구됨을 알 수 있다. 따라서, 교수·학습 과정과 평가의 통합적인 측면에서 교사의 학생평가 전문성을 분석할 필요성이 있다.

대표적인 연구로 김신영(2015)은 중등학교 교사들의 학생평가 전문성의 초점을 '학생평가의 교육적 기능 강화'에 두고 이를 위해 다음과 같은 교사의 학생평가 활동을 제안하였다.

- 수업에 앞서 학생들이 도달해야 할 성취 목표를 구체적으로 이해하고 있으며, 이를 상세화 할 수 있다.
- 학생들에게 그들이 이해할 수 있는 용어로 성취 목표에 대해 항상 알려 줄 수 있다.
- 성취 목표를 학습 정도를 확인해 낼 수 있는 평가 목표로 재진술할 수 있다.
- 학생들의 학습 동기와 평가 간의 관련성을 이해하고, 학생들의 자신감 생성을 위해 평가를 활용할 수 있다.
- 학생평가 정보를 수업 개선을 위해 지속적으로 활용할 수 있다.

- 학생들에게 자신의 학습에 대한 이해를 돕기 위해 학습 활동에 대한 피드백을 자주 제공한다.
- 학생들에게 스스로를 평가해 보는 활동에 적극적으로 참여하도록 한다.
- 학생들이 그들의 성취 수준과 학습 진전에 대해 다른 학생들과 적극적으로 소통하게 한다.
- 학생들이 그들의 현재 수준에서 도달해야 할 성취 목표가 무엇이며 다음에 해야 할 학습이 무엇인지를 기술 할 수 있게 한다.

김신영(2015)의 연구는 교사의 학생평가 전문성에 포함되어야 할 내용을 교수·학습 과정의 맥락에서 구체적으로 제시하였다는 점에서 의의가 있다.

4) 영어 교과에서의 학생평가 전문성

지금까지 학생평가 전문성의 구성요소 및 기준을 일반적인 관점에서 살펴보았다면 중등학교 영어교사들의 학생평가 전문성 평가도구를 개발하기 위해서 중등 영어교사의 학생평가 전문성의 하위 구성요소와 기준에 관한 보다 구체적인 논의가 필요할 것이다.

먼저, 다양한 해외 영어 교육 기관에서 제시하고 있는 영어교사의 평가 전문성과 관련된 내용을 정리하면 다음과 같다. QCA(Qualifications and Curriculum Authority)는 학습을 위한 평가(assessment for learning)로서 형성평가를 강조하면서 영어교사에게 필요한 능력을 다음과 같이 제시하였다(QCA 2004, 전병만 외, 2005에서 재인용).

- 학습 목표를 학생과 공유한다.
- 학생이 달성할 목표를 인식한다.
- 능력 향상 방법을 파악하는 데 도움이 되는 피드백을 제공한다.
- 모든 학생이 자신의 과거 수준보다 높은 수준까지 능력을 향상할 수 있다고 믿는다.
- 교사와 학생이 함께 학생의 수행과 발전 과정을 점검하고 성찰한다.
- 학생이 개선할 점을 스스로 발견하는 자기 평가를 할 수 있도록 한다.
- 다양한 평가 방법을 통해 효과적인 학습과 발전에 필수적인 동기 와 자신감을 북돋운다.

다음으로 TESOL(Teaching English to Speakers of Other Languages, 2002, 전병만 외, 2005에서 재인용)에서는 교사의 학생평가 전문성과 관련된 항목을 14개로 제시하고 있는데 구체적인 내용을 살펴보면 다음과 같다.

- 학습자의 수행에 관한 정보가 중요하다는 점을 인식할 수 있다.
- 교수 이전에, 교수를 하는 동안에, 그리고 교수가 끝난 후 학습자의 수행에 관한 정보를 수집하고, 해석하며, 문서로 남길 수 있다.
- 학습자의 배경, 선호, 그리고 기대에 관한 정보를 수집하고 해석할 수 있다.
- 학습자의 수행에 관한 정보를 이용하여 교수 이전에, 교수를 하는

동안에, 그리고 교수가 끝난 후 수업을 계획하고 실행할 수 있다.

- 다양한 공식적 평가도구와 비공식적 평가도구를 사용할 수 있다.
- 학습이 일어나는 과정을 모니터할 수 있다.
- 학습자들을 자기 평가에 개입시킬 수 있다.
- 평가를 학습 목표와 관련지을 수 있다.
- 다양한 방식의 체계적이고 합목적적인 평가를 사용할 수 있다.
- 학습자들이 자신의 학습을 입증할 수 있도록 하는 평가를 활용할 수 있다.
- 교수에 대한 학습자의 피드백을 활용할 수 있다.
- 교사가 제작하거나 외부에서 제작된 평가도구의 신뢰도와 타당도를 점검할 수 있다.
- 공정한 평가도구를 활용할 수 있다.
- 학습을 평가한 결과에 근거하여 학습자에게 건설적인 피드백을 제공 할 수 있다.

또한, OSE(Office for Standards in Education, 2003, 전병만 외, 2005에서 재인용)은 교사의 학생평가 전문성의 구성요소를 ① ‘수시평가를 통한 학습 과정의 모니터링’, ② ‘명확한 채점기준 사용 및 피드백 제공’ ③ ‘교수와 학습과정의 효율적 운영에 필요한 정보를 얻기 위해 평가를 활용할 수 있는 능력’ ④ ‘외부에서 설정한 준거를 활용할 수 있는 능력’의 네 가지로 제시하였다.

이상으로 살펴본 영어 교육기관에서 제시한 영어교사의 학생평가 전문성의 영역별 기준에는 다음과 같은 한계점이 발견되었다. TESOL(2002)에서 제시한 평가 전문성 기준은 교사의 학생평가와 관련된 항목을 단순히 나열하고 있을 뿐 각 구성요소 간 체계적인 관련성

은 제시하지 못하였으며, OSE(2003)와 QCA(2004)에서 제시한 영역별 기준은 주로 형성평가에만 초점을 맞춘 것으로 이를 그대로 활용하여 영어교사의 학생평가 전문성의 구성요인에 관한 개념적 모형을 제시하기에는 어려움이 존재하였다.

영어교사의 학생평가 전문성에 관한 활발한 국외 연구와는 대조적으로, 국내 연구는 많지 않은 편이다. 전병만 외(2005)의 연구에서는 영어교사의 학생평가 전문성의 구성요소를 모색하기 위하여 TESOL, OSE, QCA와 같은 영어교육 기관에서 제안하는 학생평가 전문성 기준, 일반 교육학적 관점에서 제안하는 학생평가 전문성의 구성요소와 기준 및 제7차 교육과정에서 현장 교사가 지켜야 할 평가 원칙을 제시하고 있다. 그 중 제7차 교육과정에서 필요한 교사의 학생평가 원칙을 살펴보면 다음과 같다.

- 학습한 내용을 중심으로 언어의 네 가지 기능을 균형 있게 평가한다.
- 타당도, 신뢰도, 객관도를 높이는 평가 방식을 지향한다.
- 목표에 따라 분리 평가와 통합 평가를 적절히 사용한다.
- 의사소통능력을 바르게 측정하기 위한 분리 평가보다 통합 평가의 비중을 확대해야 한다.
- 학업 성취도를 수시로 점검하여 학습 부진의 요인을 보충한다.
- 평가 결과를 교수·학습 계획에 반영한다.

제7차 교육과정에서 교사의 학생평가 원칙은 구체적인 평가 절차 및 계획에 대한 내용을 포함하고 있지 않으며 평가와 학습과의 유기적

인 연계성이 부족할 뿐만 아니라 여전히 총괄평가에 치중하고 있다는 한계점이 지적되었다.

교사의 학생평가를 교수·학습과 통합된 관점에서 살펴 본 연구로 전병만, 오준일(2006)은 학생평가 전문성의 영역별 중요성과 실행 가능성의 정도를 바탕으로 중등학교 영어과 교사의 학생평가 전문성 기준에 대한 인식을 조사하였다. 전병만, 오준일(2006)의 연구에서는 선행 연구를 바탕으로 잠정적으로 학생평가 전문성의 기준을 ① ‘평가하고자 하는 특성에 대한 이해의 정도’ ② ‘교수 및 학습 과정과 평가의 활용’ ③ ‘평가도구 개발과 활용’ ④ ‘건전한 평가에 대한 이해’ ⑤ ‘평가 결과의 산출 및 활용’의 5가지로 제시하였다. 하지만, 이 연구에서 제시된 교사의 학생평가 전문성 기준은 평가의 윤리성에 관한 고려가 부족하였다. 교사의 인성과 태도, 윤리적인 형평성과 공정성에 대해 교사가 스스로를 평가한 수준보다 학생들의 인식 정도가 더 낮은 경향이 있는 것으로 볼 때(이희경, 2007), 학생평가 전문성의 구성요소 및 기준에 있어서 윤리적인 측면이 고려되어야 함을 알 수 있다.

나. 평가 문해력(assessment literacy)의 개념 및 필요성

1) 평가 문해력의 개념

교사의 학생평가 전문성에 관한 보다 폭넓은 이해를 위해 유사한 개념인 교사의 평가 문해력에 대해 살펴볼 필요가 있다. 평가 문해력(assessment literacy)이란 교사들이 높은 수준의 실제적인 평가척도 및 채점기준을 개발하고 활용할 뿐만 아니라 학생들의 수행에 관한 올바른 판단을 하기 위해 평가와 관련된 정보를 이해, 분석, 적용할 수 있는 능력 또는 소양을 의미한다(DeLuca et al., 2016).

평가 문해력을 지닌 교사의 특성을 살펴봄으로써 평가 문해력의 개념에 대한 이해의 폭을 넓힐 수 있을 것이다. 먼저, Stiggins(1991)는 교사가 다음과 같은 다섯 가지 교실 평가의 기준을 만족시키는 방법을 습득할 때 평가 문해력을 지닌다고 하였다.

- 명확한 평가 목적을 지님
- 다양한 성취 목표 평가의 중요성을 이해
- 서로 다른 성취 목표에 따른 적절한 평가 방법의 선택
- 대표적인 수행 과제를 바탕으로 학생들의 성취에 관한 표집과 자료 수집
- 기술적 또는 실제적 문제로부터 발생할 수 있는 평가 관련 편이나 왜곡 방지

다음으로, Mertler(2005a, p.78)는 평가 문해력을 지닌 교사의 특성을 다음과 같이 요약·정리하였다.

- 학생 성취에 관한 신뢰할 만한 정보를 수집하기 위해 어떠한 평가 방법을 활용해야 할지 이해함
- 성적표, 시험 점수, 포트폴리오, 면담(conference) 등의 활용을 포함한 평가 결과에 관한 효과적인 의사소통을 수행함
- 학생들을 평가 과정에 참여시키고 활발한 의사소통을 통해 학생들의 학습 동기를 극대화하기 위해 평가를 활용함

평가 문해력을 지닌 교사들의 다양한 특성에도 불구하고 Popham(2011)은 평가 문해력을 다음과 같이 정의하면서 공통적인 세 가지 구성요소를 강조하였다.

- 모든 평가 관련 지식을 습득할 필요 없이 ①교사의 교육적 의사 결정 과정에 영향을 미칠 수 있는 ②근본적인 평가 개념 및 절차에 관한 기술적인 습득이 아닌 ③기본적인 이해 정도

다음으로 국내 연구의 경우, 김종윤(2006, 김소현, 2010에서 재인용)이 평가 문해력을 평가 전문성에 비해 보다 구체적이고 명료한 개념임을 강조하면서 그 특성을 <표 II-6>과 같이 비교하였다.

<표 II-6> 평가 전문성과 평가 문해력의 개념 비교(김중윤, 2006, p16)

	평가 전문성	평가 문해력
정의	<ul style="list-style-type: none"> • 학생 개개인의 학습과 성취에 관한 평가 정보를 수집하고 이를 활용할 수 있는 능력(김신영, 2004; 이인제 외, 2004, p.27) 	<ul style="list-style-type: none"> • 평가에 대한 명세적·방법적 지식으로 이 지식은 교육적 목적에 맞추어 건전한 평가를 이해, 선택하고, 이러한 평가를 개발 및 실행할 수 있으며, 평가의 결과를 적절하게 해석하여 다른 사람과 의사소통할 수 있는 지식을 포함(Stiggins,1995; Mertler,2004 등)
장점	<ul style="list-style-type: none"> • 교사의 평가 능력을 포괄적으로 살펴볼 수 있음(교육과정, 교과내용, 교과방법에 대한 전문적 지식을 요구) 	<ul style="list-style-type: none"> • 평가에 대한 능력(지식, 기능, 태도)에 초점을 맞추어 살펴볼 수 있음 • 개념 정의가 비교적 명료함
단점	<ul style="list-style-type: none"> • 너무 광범위한 개념으로, 교사의 전문성 일반과 혼동이 됨 	<ul style="list-style-type: none"> • 평가에 대해 초점을 맞추어 평가자로서 중요하게 다룰 평가 외적인 요소 (교육과정, 교과내용, 교과방법 등)를 소홀히 할 우려가 있음.
특징	<ul style="list-style-type: none"> • 평가 문해력을 포함하는 광의의 개념 • 교과에 대한 배경 지식 및 평가에 대한 지식을 갖추어야 평가 전문가로 볼 수 있다는 관점 	<ul style="list-style-type: none"> • 평가 전문성에 포함되는 협의의 개념 • 평가에 대한 지식(평가 문해력)을 바탕으로 교과 상황에서 활용해야 평가 전문가로 볼 수 있다는 관점

즉, 평가 문해력은 주로 교사의 학생평가 활동에 초점을 맞춘 개념으로 보다 좁은 의미의 평가 전문성이라고 할 수 있다.

2) 평가 문해력의 필요성

교사는 21세기에 필수적인 역량을 학생들에게 가르치고 평가하기 위해 필요한 지식 및 기술을 숙달해야 한다는 점에서 평가 문해력 신장을 위한 교사의 전문성 개발 노력이 더욱 강조되고 있다. 김소현(2010)은 평가 문해력이 교사의 평가 책무성에 대한 인식 제고에서 시작된다고 언급하면서, 특히, 근접발달영역(ZPD)에서 교사의 전반적인 평가 문해력이 학생들의 학업 성취에 결정적인 역할을 한다고 지적하였다. Popham(2011)은 교사의 평가 책무성이 교육의 질을 결정하며, 교수 활동과 통합된 평가로서의 형성평가가 강조되면서 교사의 평가 문해력은 학생들의 학습의 질을 향상시키기 위한 수업 관련 의사 결정에 있어서도 큰 역할을 한다고 주장하였다.

지난 20년 동안, 실제적 평가(authentic assessment)를 위한 교육 개혁은 전 세계적으로 활발하게 이루어졌으나, 교사들의 평가 문해력(assessment literacy)은 그 중요성에도 불구하고 제대로 향상되지 못하였다. 교사들은 실제 자신의 평가 활동에 관해 갖는 자신감이 부족하거나 평가 문해력을 신장시키기 위한 다양한 자료나 연수 기회가 상대적으로 부족하였다(주세형, 2011; Kiomrs et al, 2011; Koh, 2011; Popham, 2011; Xu & Brown, 2016).

교사의 평가 문해력 신장을 위해서는 단기간의 평가 연수나 일회성의 워크숍 참여보다는 장기간의 사전 연수(pre-service teacher training programs)나 현직 연수 프로그램(in-service teacher training programs)의 참여 등 지속적인 전문성 개발 노력이 더욱 효과적인 것으로 나타났다(McMunn et al., 2004; Wiliam & Thompson, 2008;

Koh, 2011). 특히, Wiliam와 Thompson(2008)은 학교 단위의 활발한 교사 협의회 활동 참여가 교사에게 학생평가 활동에 관한 자기반성의 기회를 부여하며 동료 교사의 도움으로 문제를 해결할 수 있다는 점에서 동료 교사와의 협력을 통한 전문성 개발의 필요성을 강조하였다.

다음으로, Koh(2011)는 싱가포르 교사를 대상으로 한 종단 연구에서 효과적인 교사의 평가 전문성 개발 프로그램을 위해 내용 중심(content focus), 능동적 학습(active learning), 일관성(coherence), 기간(duration)과 집단적 참여(collective participation)를 주요 특징으로 강조하였다.

또한, Xu와 Brown(2016)은 교육평가와 교사교육 측면에서 평가 문해력을 살펴볼 필요가 있다고 지적하면서 평가 문해력을 신장시키기 위한 노력을 교사 개인적 차원, 기관적 차원, 국가적 차원으로 구분하여 구체적으로 제시하였다. 개인적 차원의 노력으로 교사는 자신의 전문적 정체성을 교수자(instructor)에서 학습 평가자(assessor)로 재개념화할 필요가 있다. 다음으로, 기관적 차원의 노력의 일환으로 교사의 학생평가 활동에 영향을 주는 학교에서의 권력관계와 같은 구조적 조건을 개선하는 것이다. 마지막으로 국가적 차원에서 교사 전문성 개발 활동이나 교사의 자율성을 신장시킬 수 있는 구체적인 다양한 정책적 마련을 제시하였다.

이와 같은 평가 문해력 신장을 위한 노력은 교사의 학생평가 전문성 향상에 있어서도 의미 있는 시사점을 제공할 것이다. 이 연구에서는 교사의 학생평가 전문성을 앞서 살펴본 교사의 평가 문해력의 개념에 평가의 외적 요소(교육과정, 교과내용, 교과 방법 등)에 관한 전문적 지식을 포함하는 보다 포괄적인 개념으로 이해하여 기술하였다.

다. 교사의 학생평가 전문성 함양 실태

최근 교실 내 학생평가 활동에서 학생의 학습에 대한 이해 및 학습 활동의 개선을 지원하고 학생들 스스로가 학습에 대한 흥미와 보람을 느낄 수 있도록 학생평가의 교육적 기능이 강조되고 있다. 따라서, 학생평가 활동의 교육적 기능을 보다 향상시키기 위해서는 학교 현장에서 교사의 학생평가 전문성 함양 실태에 관한 정확한 파악이 선행되어야 할 것이다.

교사의 학생평가 전문성 함양에 관한 실행 수준을 구체적으로 살펴보면, 학생평가 전문성의 높은 교육적 요구에도 불구하고, 많은 교사들은 스스로 학생평가 전문성 향상을 위해 개인적인 노력을 기울이지 않거나 (Mertler, 2005b), 학생평가 전문성 함양을 위한 교육 기회를 적절하게 제공받지 못한 것으로 나타났다(김신영, 2006; 배호순, 2008; 전병만 외, 2005).

‘교사의 학생평가 실태 조사 및 전문성 신장에 대한 요구 분석(이인제 외, 2004)’에 따르면 우리나라의 경우, 예비 교사들을 위한 교사 양성교육 기관 및 현직 교사 연수의 교육 내용과 방법, 프로그램이 교사의 학생평가 전문성을 신장시키기 위한 체계적인 시스템을 갖추지 못하였다. 또한, 교육 측정 및 평가 이론이 교사의 실제적인 학생평가 전문성 함양에 큰 도움을 주고 있지 못하는 것으로 나타났다(김신영, 2002; 김신영, 2007). 이는 평가 관련 교사 연수에서 중점적으로 다루고 있는 내용이 평가도구의 개발 또는 학습 결과에 대한 평가에만 국한되어 교수·학습 과정 속에 통합된 새로운 학생평가의 개념을 제대로 반영하지 못한 결과라고 할 수 있다.

김신영(2015)은 교사 대상 심층 면접과 설문조사를 통해 ① ‘학교의

평가 계획과 실천' ② '진단·형성·총합평가의 운영 실태' ③ '수행평가' ④ '성취평가제' ⑤ '학교생활기록부 기재' 의 다섯 가지 측면에서 학교 현장에서의 학생평가의 운영 실태와 교사의 학생평가 활동의 실행 수준을 살펴보았다. 분석 결과, 학생평가 활동의 특성에 관한 교사들의 전반적인 이해 수준은 높았지만 구체적인 실행 방법에 대한 이해 수준은 상대적으로 낮은 것으로 나타났다.

또한, 박지현 외(2017)는 2015 개정교육과정에서 과정중심평가가 강조되고 수행평가의 유형이 다양화되고 있는 현재의 교육적 맥락에서 단위학교에서 수행되는 학생평가의 운영 실태와 이해관계자들의 인식을 분석하였다. 박지현 외(2017)의 연구에서는 교사의 학생평가 운영 실태를 '학생평가 유형 및 교육과정 연계', '채점 계획 및 질 관리', '평가 결과의 활용 및 피드백', '교사의 평가전문성 개발' 의 네 가지 측면에서 분석하였으며 주요 연구 결과를 정리하면 다음과 같다. 첫째, 총합평가에서 벗어나 다양한 유형의 평가가 활용되고는 있지만 여전히 기존의 평가방식을 그대로 활용하는 경우가 많았다. 둘째, 수행평가와 서술형 평가를 강화하는 정책이 시행되면서 채점의 공정성을 확보하고자 노력하고 있으나 채점 결과의 신뢰도 확보를 위한 구체적인 방안이 마련되지 않은 실정이다. 따라서, 채점의 공정성을 확보하고 교실 내 평가 활동의 효과를 향상시키기 위해 동료 교사 간의 협업이 필요할 것이다 (손원숙 외, 2015). 셋째, 교사 및 학교 차원에서 교수·학습 방법 개선을 위해 피드백을 활용하고 있으나 학생들은 평가 결과에 대한 교사의 구체적인 피드백 제공이 부족하다는 인식이 많았다. 넷째, 교사의 학생평가 전문성 개발 측면에서 살펴보면 많은 교사들이 학생평가 실행에 대한 자신감이 낮았으며 평가관련 교사 연수의 실효성에 부정적인 인식을 지니고 있는 것으로 나타났다.

이와 마찬가지로, 중등 영어교사들의 학생평가 전문성 함양 실태를 살펴보면, 중등 교사 양성을 위한 사범대학이나 교육대학원의 교과 과정에서 영어 평가와 관련된 교과가 필수 과목으로 지정되지 않거나(전병만 외, 2005), 현직 중등 영어교사들을 위한 연수 프로그램이 지나치게 영어 교수법이나 기술적인 지식에만 초점을 두어 실제적인 평가 전문성을 함양할 기회가 절대적으로 부족한 것으로 지적되었다(김진완 외, 2004).

따라서, 교사의 학생평가 전문성이 어느 때보다 요구되고 있는 현 시점에서 교수·학습의 효과를 확인하고 교육 과정 전반의 효율적인 운영을 위해서 교원 양성 및 연수 기관의 교육 과정 운영, 교육 내용과 방법, 연수 프로그램의 이수 기회 및 시기, 연수 내용과 방법 등에 대한 체계적인 점검이 필요하다(김경희 외, 2006). 즉, 교수·학습 과정에 통합된 개념으로서의 학생평가 활동을 이해하고 교사의 평가 전문성을 향상시키기 위한 다양한 교원 양성 과정 및 현직 교사 연수 프로그램이 개발되어야 할 것이며, 이를 학교 현장에서 적용하기 위한 행정적, 재정적 지원체제가 수반되어야 할 것이다(이인제 외, 2004).

2. 교사의 학생평가 전문성 평가도구

교사의 평가 전문성에 대한 관심이 높아지면서 교사의 학생평가 전문성을 평가할 수 있는 도구 개발에 대한 필요성 또한 증가하고 있다. 중등학교 영어교사의 학생평가 전문성 평가도구 개발에 앞서 보다 일반적인 측면에서 교사의 학생평가 전문성을 측정하기 위한 평가도구와 관련

된 선행 연구를 분석해 볼 필요가 있다. 관련 선행 연구는 주로 평가하고자 하는 학생평가 전문성의 내용적 측면과 방법론적 측면을 중심으로 분류할 수 있다.

가. 내용적 측면

내용 영역을 중심으로 학생평가 전문성과 관련된 국외 선행연구를 살펴보면 다음과 같다. AFT 외(1990)의 기준에서 제시한 7개의 평가 전문성 구성요소를 바탕으로 한 ‘Teacher Assessment Literacy Questionnaire(TALQ)’ (Plake et al., 1993), ‘Assessment Practices Inventory(API)’ (Zhang & Burry-stock, 1997), ‘Assessment Literacy Inventory(ALI)’ (Campbell et al., 2002), ‘Classroom Assessment Literacy Inventory(CALI)’ (Mertler, 2003), 그리고 TALQ를 수정·보완한 ‘Revised ALI’ (Mertler & Campbell, 2005) 등이 예비 교사 및 현직 교사의 평가 전문성을 측정하기 위해 개발되었다.

그 외의 평가도구로는 Daniel 과 King(1998)이 30개의 이분 문항(True/False)으로 교사의 평가 전문성을 측정한 ‘Measurement Literacy Questionnaire’ 와 Jarr(2012)이 학생평가 활동과 관련된 교사 효능감 정도를 측정하기 위해 개발한 7점 척도의 15문항으로 구성된 ‘Assessment Self-Confidence Survey’ 가 있다.

다음으로, 국내 선행연구의 경우, 교사의 학생평가 전문성 평가도구 개발에 앞서 보다 포괄적인 측면에서 교사 전문성의 인지적인 영역을 측정한 연구(안미화, 2005; 이경아, 김경희, 2006; 최윤영 외, 2012)와 정의적 영역을 측정한 연구(임수연, 2014; 최진령, 이연승, 2010;

Swailers, 2003)로 구분할 수 있다.

특히, 임수연(2014)은 교사 전문성을 정의적 관점에서 측정할 수 있는 평가도구 개발의 필요성에 따라 중등학교 교사와는 차별적인 특성을 지닌 초등학교 영어교사를 대상으로 ① ‘소명의식’ ② ‘전문적 자율성’ ③ ‘공공 봉사심’ ④ ‘전문적 준거성’ ⑤ ‘자기 통제에 대한 믿음’의 다섯 가지 요인으로 교사 전문성을 측정하고자 하였다. 임수연(2014)의 연구는 교사 전문성이 교실 내의 교수·학습 과정 뿐만 아니라 교실 밖의 요인인 조직적, 문화적, 교사 공동체와 밀접한 관계를 지님을 밝히면서 다른 교사와의 협업 능력과 같은 공동체적 관점에서 교사 전문성을 측정해야함을 시사하였다. 임수연(2014)이 개발한 ‘초등학교 영어교사의 전문성 평가도구’의 경우 현직 초등학교 영어교사로 구성된 면담그룹과의 논의를 통해 초등학교 영어교육의 상황에 적합한 평가도구를 개발했다는 점에서 의의가 있지만 측정 대상을 중·고등학교 교사로 확대하거나 진단 결과를 교사의 학생평가 전문성으로 일반화하기에는 한계점이 발견되었다.

다음으로, 교사 전문성의 인지적, 정의적 영역을 모두 고려한 연구로는 델파이 방법을 활용하여 윤리 교사의 전문성을 측정할 수 있는 평가도구를 개발한 김길모 외(2014)의 연구가 있다. 이 연구에서는 윤리 담당 교사의 전문성 구성요소를 ① ‘정보윤리 지식’ ② ‘정보윤리 교육학 지식’ ③ ‘도덕적 판단능력’ ④ ‘학생상담 및 대응 능력’ ⑤ ‘학급운영 및 수업 능력’ ⑥ ‘인식’ ⑦ ‘태도’의 7가지로 도출하였다.

이들 연구는 공통적으로 교사의 수업 기술과 지식의 인지적 요소 뿐만 아니라 신념, 소명 의식, 헌신도, 전문적 자율성과 같은 정의적인 요소가 전반적인 교수·학습 활동에 미치는 영향에 관해 측정하였다는 점에서 의의를 지니나, 교사의 학생평가 전문성을 구체적으로 측정하지는

못하였다.

교사의 학생평가 전문성 평가에 초점을 두고 개발된 대표적인 연구로는 한국교육과정평가원에서 수행된 송미영, 김경희(2007)의 연구가 있다. 이 연구에서는 교사의 학생평가 활동에 필수적인 평가 지식과 기능을 중점적으로 교사 스스로 자신의 평가 전문성을 진단할 수 있도록 ‘교사의 학생평가 전문성 진단도구’를 개발하였다. 이 평가도구는 전문가 집단의 합의와 타당화 연구를 거쳐 설정된 교사의 학생평가 전문성에 대한 절대적인 기준선(baseline)을 제시함으로써 교사 스스로 부족한 평가 능력의 요소와 수준을 파악할 수 있도록 하였다. ‘교사의 학생평가 전문성 기준’은 평가 관련 교원 양성 프로그램 및 현직 교사를 위한 연수 프로그램의 효과를 측정하고 연수에 참여한 교사들의 평가 전문성 향상 정도를 측정하기 위한 도구로도 활용될 수 있다(김수동 외, 2005b). 또한, 최지선(2014)은 해당 평가도구를 활용하여 교사의 학생평가 전문성 중 특히 ‘평가도구 개발 영역’에 주목하였으며, 현직 수학 교사를 대상으로 평가 문항 개발 연수가 교사들의 평가도구 개발 전문성 향상에 효과가 있는지를 분석하였다.

이와 같이 ‘교사의 학생평가 전문성 진단 도구’는 교사의 평가 전문성에 대한 절대적인 기준선을 제공함으로써 교사들에게 자기 장학의 기회를 제공하고 기준에 도달하지 못한 교사에게는 교사 연수의 준거로서의 역할을 한다는 점에서 의의가 있다. 하지만, 주된 내용이 학생평가 활동을 위해 교사가 알고 있어야 하는 ‘지식’ 및 ‘기능’과 같은 평가전문성의 인지적, 행동적 측면만을 강조함으로써 정의적 영역을 포함한 교사의 학생평가 전문성을 전반적으로 이해하고 측정하는 데에는 한계점이 발견되었다(정훈, 2011; Shin, 2012).

나. 방법론적 측면

다음으로, 국내·외의 선행연구를 바탕으로 교사의 학생평가 전문성 평가도구를 방법론적 측면에서 분석하면 ①자기보고식 설문조사 또는 비구조화된 면담을 활용한 ‘자기 평가 방법’ (권순달, 2010; 김선희, 2006; 김선희, 2012; 김신영, 2002; 김신영, 2007; 김신영, 송미영, 2008; 박가나, 2012; 박정, 2013; 전병만, 오준일, 2006) 과 ②교사 포트폴리오를 활용한 ‘전문가 평가 방법’ (윤상이, 2015; Webb, 2012)으로 유형화 할 수 있다.

자기보고식 검사를 활용한 자기 평가 방법은 데이터 수집과 결과 분석이 비교적 용이하다는 장점이 있으나, 피검사자가 사회적으로 바람직한 방향으로 자신을 왜곡하는 ‘긍정적 반응왜곡(faking-good)’ 으로 인해 검사 결과에 대한 신뢰도와 타당도 저하의 문제점이 발생하며(어윤경, 2006), 실제 수행에 근거한 교사 전문성을 일회성 또는 단기간의 설문조사로 평가하기에는 어려움이 따른다. 실제로, 중등교사들의 실천적 교수역량에 대한 자기 평가와 전문가 평가의 결과를 비교한 황은희, 백순근(2008)의 연구에서 교사들은 스스로에 대해 보다 관대한 평가를 하는 경향이 있음 볼 때, 교사의 학생평가 전문성을 보다 정확하고 객관적으로 평가하기 위해서는 교사 포트폴리오를 통한 전문가 평가 방법을 활용할 필요가 있다.

교사 포트폴리오(teaching portfolio)는 “교사 자신의 교육 활동에 대한 성찰을 통해 향상과 발달이 이루어질 수 있도록 학생의 학습과 발달에 영향을 미치는 교사의 교수 활동의 실제, 수업과 관련된 각종 활동과 결과물, 이에 대한 반성적 진술과 설명 등이 포함된 조직적인 기록” 이다(손영민, 2007, P182). 즉, 교사 포트폴리오 평가는 교사의 실천적

지식을 향상시키며 반성적 사고를 증진시키는데 효과적인 교사 교육 및 교사 평가에 유용한 방법이라 할 수 있다(김혜영, 2004; 황해익, 2000, 안진경, 2011에서 재인용). 교사 포트폴리오의 내용으로는 교육 철학, 교수·학습 자료, 수업지도안, 수업을 관찰한 일지 내용, 수업을 녹화한 비디오테이프, 학습자 피드백, 학습 결과, 수업에 대한 교사 자신의 반성과 소감 등이 포함될 수 있다(Baily, Curtis & Nunan 2001; Van der Westhuizen & Smith, 2000, 전병만 외, 2006에서 재인용).

교사 포트폴리오를 활용한 평가는 다음과 같은 장점을 지닌다. 첫째, 일회성 관찰로 이루어진 수업 평가와는 대조적으로 장기간의 교사의 교육 철학 및 경험, 의사결정 과정, 성찰(reflection), 자격, 전문성 등 눈에 보이지 않는 교사 개인의 특성이 잘 드러날 뿐만 아니라 교수 활동이 이루어지는 문맥에 바탕을 둔 보다 전반적인 교사 평가가 가능하다. 둘째, 포트폴리오의 특성상 누적적 자료 수집과 활용이 용이하여 지속적으로 교수·학습 과정에서 발견되는 문제점을 진단하고 개선책을 마련할 기회를 제공할 수 있다(김혜영, 2004; 손영민, 2007; 안진경, 2011; 임승렬, 2000; 전병만 외, 2006; Shulman, 1998). 셋째, 교수 활동의 기본 바탕과 교사의 행동을 형성하는 의사결정 능력을 발달시키고 다양한 교육적 통찰력을 다른 사람과 공유하는 등 교사 교육을 통한 교사 전문성 신장에 있어서 매우 유용하다. 넷째, 교사들 간의 토론을 통하여 교수 활동에 관한 의견을 교환하는 등 협력적인 교사 문화의 형성에 기여할 수 있다(손영민, 2007).

반면에, 포트폴리오 개발 및 구성에 비교적 많은 시간을 할애하거나 교사 스스로 자신의 교수 활동의 자료를 선별적으로 수집하는 것이 어려운 점, 교사 스스로 자신의 포트폴리오를 평가할 경우 나타나는 측정의 신뢰도 문제 등은 교사 포트폴리오를 활용한 평가의 한계점으로 지적할

수 있다(임승렬, 2000).

교사 포트폴리오를 활용한 선행 연구를 살펴보면, 주로 비디오 포트폴리오를 활용하여 교사의 실천적 교수역량을 평가한 연구(윤승혜, 2011; 이재열, 2008; 함은혜, 2007; 황은희, 2008), 웹기반 교수 포트폴리오를 활용하여 영어교사 평가 모형을 제시한 전병만 외(2006)의 연구, 예비 교사들을 대상으로 학습 포트폴리오 활동의 효과에 대한 연구(손유진, 황해익, 2004; 안진경, 2011), 평가 포트폴리오를 활용하여 학생평가 전문성을 측정한 연구(윤상이, 2015; Webb, 2012) 등이 있다.

특히, 윤승혜(2011), 이재열(2008), 함은혜(2007), 황은희(2008)의 연구에서는 교사의 실천적 교수 역량을 평가하기 위한 목적으로 교사의 수업 장면을 녹화하여 포트폴리오를 작성 한 후 전문가들이 평가하는 비디오 포트폴리오 방법을 활용하였다. 비디오 포트폴리오 평가는 교수·학습 과정에서의 상호 작용을 쉽게 파악할 수 있으며 교사 자신 및 동료 교사의 교수 활동에서의 장점과 개선할 점을 쉽게 파악할 수 있어 교사의 실천적 교수 역량을 평가하는데 유용한 방식이다(백순근, 1999).

한편, 교사의 학생평가 전문성을 평가하는데 있어서 교사 포트폴리오를 활용한 연구는 많지 않았다. 국외 연구의 예로 Webb(2012)은 32명의 수학 교사들을 대상으로 한 교사 연수 프로그램에서 평가 포트폴리오(assessment portfolio)를 활용하여 2년 동안 수학 교사들의 평가 전문성의 변화를 종단적으로 살펴보았다. 이때 평가 포트폴리오는 ‘학생들의 학습에 관한 교사의 의사 결정을 위해 사용된 모든 평가 과제’를 의미한다. Webb(2012)의 연구는 수집된 포트폴리오에 대한 3명의 전문가들의 평가를 통해 ①평가 과제 선택 및 실시에 관한 ‘평가 계획’ ② ‘평가 결과의 해석’ ③학생들의 이해 정도 및 학습 개선에 도움을 주는

사고 능력을 평가하는 ‘평가 전략의 제시’ ④ ‘동료 교사와의 토론 및 협업’ 능력 측면에서 교사들의 실제적인(authentic)평가 전문성에 관한 경험적 자료를 제시하였다.

국내 연구의 경우, 초등학교 교사들의 학생평가 전문성을 측정하기 위해 평가 포트폴리오를 바탕으로 전문가 평가 방법을 활용한 윤상이(2015)의 연구가 있다. 이 연구에서는 초등학교 교사의 학생평가 전문성 척도와 초등학생들의 수업 만족도 측정 도구를 각각 개발하여 교사의 학생평가 전문성이 학생들의 수업 만족도에 미치는 영향을 분석하였다. 윤상이(2015)의 연구에서 개발된 초등학교 교사의 학생평가 전문성 척도의 구성요소와 하위 문항을 구체적으로 살펴보면 <표 II-7>과 같다.

<표 II-7> 초등학교 교사의 학생평가 전문성 척도(윤상이, 2015)

구성 요인	내용	문항
평가 계획	학생의 학습목표 도달여부에 대한 정보를 얻기 위해 구체적이고 타당한 평가 계획을 세우는 능력	1.[평가안내]평가 계획 및 방법, 채점기준에 대한 안내를 사전에 충분히 하는가?
		2.[평가목표]학생이 달성해야 할 학습목표를 바탕으로 하여 평가목표를 명확하게 제시하는가?
		3.[평가내용]평가계획에 평가방법, 성취기준, 성취수준을 포함하여 제시하는가?
		4.[평가방법]평가방법은 교수·학습 과정에서의 학습 활동을 반영 또는 응용한 것인가?
		5.[평가방법]적절한 학습방법을 유도할 수 있는 평가방법인지 고려하는가?
		6.[평가참여]학생에게 자기평가 및 동료평가 등 평가에 참여 할 기회를 제공하는가?
		7.[평가검토]평가목적과 용도에 적합한 평가인지 검토 할 수 있는가?
평가도구 선정 및 개발	평가목표에 적합한 평가방법을 선정 및 개발하는 정도	8.[학생수준]학생의 수준에 적합한 과업 및 문항인가?
		9.[학생흥미]학생의 흥미를 고려한 과업 및 문항가?
		10.[수행맥락]과제를 수행하는 맥락이나 상황을 제시 하는가?
		11.[채점기준]채점기준을 명확하게 제시하는가?

		12. [개발능력] 평가의 목적과 내용에 부합하는 평가 도구가 없을 경우 평가도구를 직접 개발 할 수 있는가?
		13. [도구이해] 평가도구의 특징 및 제한점을 이해하고 있는가?
실행 및 분석	평가를 실행하는 과정에서 요구되는 채점, 결과 해석, 성적 부여 능력	14. [평가시간] 평가 시간은 적절한가?
		15. [수행과정] 학생들에게 과제 수행이 복잡하지 않은가?
		16. [채점과정] 채점 과정은 복잡하지 않은가?
		17. [결과제시] 평가내용 및 방법을 고려했을 때 평가 결과 제시 방식 (점수, 등급, 서술 등)은 적절한가?
		18. [결과해석] 평가 결과에 기초하여 성취도를 판단할 수 있는가?
		19. [결과해석] 학생의 성취를 종합적으로 이해하기 위해 다양한 양적, 질적 자료와 정보를 수집 또는 기록하는가?
결과 활용	평가의 결과를 학생에게 피드백하고 교수·학습과정에 활용하는 능력	20. [피드백] 학생들에게 평가 결과에 대한 피드백을 제공하는가?
		21. [피드백] 피드백 내용은 개선 방향을 제시하고 있는가?
		22. [교수학습] 학생들에게 자신의 과제 수행에서 나타난 문제점을 분석하고 개선할 수 있는 기회를 제공하는가?
		23. [교수학습] 평가결과에서 얻은 정보를 교수·학습 자료로 재활용 하는가?
		24. [교수학습] 평가결과를 토대로 수업 내용의 양과 수준(보충·심화 등)을 조절하는가?
		25. [의사소통] 학생들의 학습활동, 평가 계획 및 결과에 대한 정보를 학생의 보호자와 공유하는가?

윤상이(2015)의 연구는 ‘서술형 설문지’, ‘수행평가 계획서’, ‘수행평가 문제지’와 ‘채점된 답안지’, ‘학생평가 자료’ 및 ‘기록물’ 등 교사의 학생평가 활동과 관련된 교사 포트폴리오를 활용한 전문가 평가를 실시하였다. 이 연구는 기존의 자기보고식 설문조사를 통한 자기 평가 방법에 비해 평가의 실체성 및 초등학교에서의 현장 적합성을 높여 초등교사의 학생평가 전문성 평가를 위한 중요한 기초 자료를 제공한 점에서 의의를 지닌다.

하지만, 초등교사의 학생평가 전문성 척도 개발과정에 있어서 ‘평가 계획’, ‘평가도구의 선정 및 개발’, ‘평가 실행 및 분석’ 과 ‘결과 활용’ 의 4가지 구성 영역만을 포함하여 ‘평가의 윤리성’ 에 관한 고려가 부족한 점, 학생평가 전문성의 구성 요인 및 기준에 대한 이론적 준거가 명확하지 않은 점, 평정 기준이 모호하여 채점자 간 신뢰도에서 문제점이 발견된 점은 이 연구의 한계점으로 지적할 수 있다. 또한, 연구의 주된 관심을 교사의 학생평가 전문성이 학생들의 수업 만족도에 미치는 영향에 둬으로써 평가 전문성 척도 자체의 문항 타당도 및 신뢰도에 대한 체계적인 검증이 이루어지지 않았다. 따라서, 교사의 학생평가 전문성에 대한 폭넓은 이해를 바탕으로 한 이론적 모형을 재설정하고 타당화 연구를 통해 척도의 타당도 및 신뢰성을 확보한 새로운 평가도구의 개발이 요구된다.

이상으로 살펴본 선행연구를 바탕으로 교사의 학생평가 전문성 평가도구의 하위 영역 및 측정 방식을 정리하면 <표 II-8>과 같다.

〈표 II-8〉 교사의 학생평가 전문성 평가도구의 하위 영역 및 측정 방식

선행연구	하위 영역	측정 방식
<ul style="list-style-type: none"> • Plake et al.(1993) • Zhang&Burry-stock (1997) • Campbell et al.(2002) • Mertler (2003) • Mertler&Campbell(2005) 	①평가 방법의 선정	자기평가방법 (자기보고식 설문조사)
	②평가도구의 개발	
	③평가의 실시 및 채점과 결과 해석	
	④평가 결과의 활용	
	⑤성적부여	
	⑥평가 결과에 관한 의사소통	
	⑦평가의 윤리적 측면에 대한 인식	
<ul style="list-style-type: none"> • Daniel & King(1998) 	①교육 측정· 평가 관련 지식	자기평가방법 (자기보고식 설문조사)
	②평가 결과의 해석	
	③평가 결과에 관한 의사소통	
<ul style="list-style-type: none"> • 이인제 외 (2004) 	①평가 방법의 선정	자기평가방법 (자기보고식 설문조사)
	②평가도구의 개발	
	③평가의 실시 및 결과 분석	
	④평가 결과의 보고	
	⑤평가 결과의 활용	
	⑥윤리적 문제의 인식	
<ul style="list-style-type: none"> • 김경희 외(2006) • 김신영 (2007) • 송미영, 김경희 (2007) • 김신영, 송미영 (2008) • 박가나 (2012) • 박정 (2013) 	①평가방법의 선정	자기평가방법 (자기보고식 설문조사)
	②평가도구의 개발	
	③평가실시·채점·성적부여	
	④평가결과의 분석·해석·활용·의사소통	
	⑤평가의 윤리성	
<ul style="list-style-type: none"> • 김선희 (2006) 	①평가방법의 선정	반구조화된 면담 (중등교사 11명 대상)
	②평가도구의 개발	
	③평가실시·채점·성적부여	
	④평가결과의 분석·해석·활용	
	⑤평가의 윤리성 인식	
<ul style="list-style-type: none"> • 전병만, 오준일 (2006) 	①평가하고자 하는 특성에 대한 이해 정도	자기평가방법 (자기보고식 설문조사)
	②교수 및 학습 과정과 평가의 활용	
	③평가도구 개발과 활용	
	④건전한 평가에 대한 이해,	
	⑤평가결과의 산출 및 활용	
<ul style="list-style-type: none"> • 이희경 (2007) 	①평가운용능력	자기평가방법 (자기보고식 설문조사)
	②지필평가 관련 문항제작능력	
	③수행평가 능력	
	④평가결과의 표수 반영 실태	
	⑤윤리적인 형평성 및 공정성	

• 김소현 (2010)	지식	①평가의 목적이해 ②하위 평가도구의 이해 ③학습자에 대한 이해 ④평가 과제 이해	
	수행	①평가의 설계 ②성취기준, 평가 기준, 과제 개발 ③평가의 적용 ④평가 결과의 해석 및 활용	자기평가방법 (자기보고식 설문조사)
	태도	①평가에 대한 성실한 태도 ②평가에 대한 공정한 태도 ③평가 전문성 계발 노력	
• 김선희 (2012)		①평가방법의 선정 ②평가도구의 개발 ③평가 실시·채점·성적부여 ④평가 결과의 분석·해석·활용·의사소통 ⑤평가의 윤리성	①양적분석방법: 자기보고식 설문조사
		①평가 관련 지식: 개념적 지식 및 기술적 지식 ②학생과 교사의 이중적 입장: 평가의 윤리적 측면 및 학교 현장과 평가 이론의 괴리 ③평가 실행: 평가도구개발, 평가실시, 채점, 결과 분석,	②질적분석방법: 예비교사의 포트폴리오(학습일지) 분석
• Jarr (2012)		①표준화된 평가 결과 해석 ②평가 결과의 활용 ③평가 결과에 관한 의사소통 ④평가의 법적·윤리적 측면에 대한 인식	자기평가방법 (자기보고식설문조사)
• Webb(2012)		①평가 계획 ②평가 결과의 해석 ③평가 전략의 제시 ④동료교사와의 토론 및 협업능력	평가포트폴리오 (assessment portfolio)를 활용한 전문가 평가
• 최지선(2014)		‘교사의 학생평가 전문성 진단 도구(송미영, 김경희, 2007) 중 ‘평가도구의 개발’ 영역	자기평가방법 (자기보고식 설문조사)
• 윤상이(2015)		①평가 계획 ②평가도구 선정 및 개발 ③평가 실행 및 분석 ④결과 활용	평가포트폴리오 (assessment portfolio)를 활용한 전문가 평가

다. 선행 연구의 한계점

교사의 학생평가 전문성 평가도구에 관한 다양한 국외 선행연구에 비해 국내 연구의 경우, 한국교육과정평가원에서 수행한 송미영, 김경희(2007)의 ‘교사의 학생평가 전문성 진단도구’를 활용한 연구가 대부분이었다. 이상 교사의 학생평가 전문성 평가도구와 관련된 국내 선행연구에서 공통적으로 발견된 한계점을 정리하면 다음과 같다.

첫째, 교육 과정의 내적 구조 틀에서의 평가에만 초점을 두어 교사의 사회적, 정치적 영향력 및 맥락 인식 능력을 형성하려는 노력에 관한 논의가 부족하였다(김혜정, 2008). 둘째, 주로 교사 평가 전문성의 구성 요인이나 성취 지표에만 관심을 둘 뿐 전문성 신장을 위한 구체적인 교육 내용은 제시하지 못하였다(주세형, 2011). 셋째, 교사의 학생평가 전문성의 인지적 측면에만 관심을 둘 뿐 평가의 윤리성 외에 교사가 평가에 대해 갖는 전반적인 관점이나 태도 등의 평가 전문성의 정의적 요소에 관한 고려가 부족하였다(임수연, 2014). 넷째, 교과 영역별 평가 기준 연구가 수행되기 하였으나 애초에 특정 교과를 위한 평가 전문성 개발 기준이 존재하지 않아 실제로 특정 교과의 교사가 활용할 수 있는 기준으로서의 적합성을 기대하기 어려운 측면이 있다. 평가 전문성과 같은 교사의 잠재적 실천역량은 상황 또는 영역 특수성(domain specificity)을 지님을 볼 때(백순근, 임현수, 2006), 교사의 학생평가 전문성 평가도구를 모든 교과목의 교사들에게 일괄적으로 적용하는 것은 적절하지 않을 것이다. 따라서, 평가 전문성의 과목 특수적, 맥락 의존적 특성을 고려하여 과목 특성과 평가 대상을 보다 구체화한 평가도구 제작의 필요성이 제기된다(Xu & Brown, 2016). 다섯째, 주로 교사 개인의 관점에서만 학생평가를 인식하고 있어 평가의 주체로서 학생, 동료교사 및 교

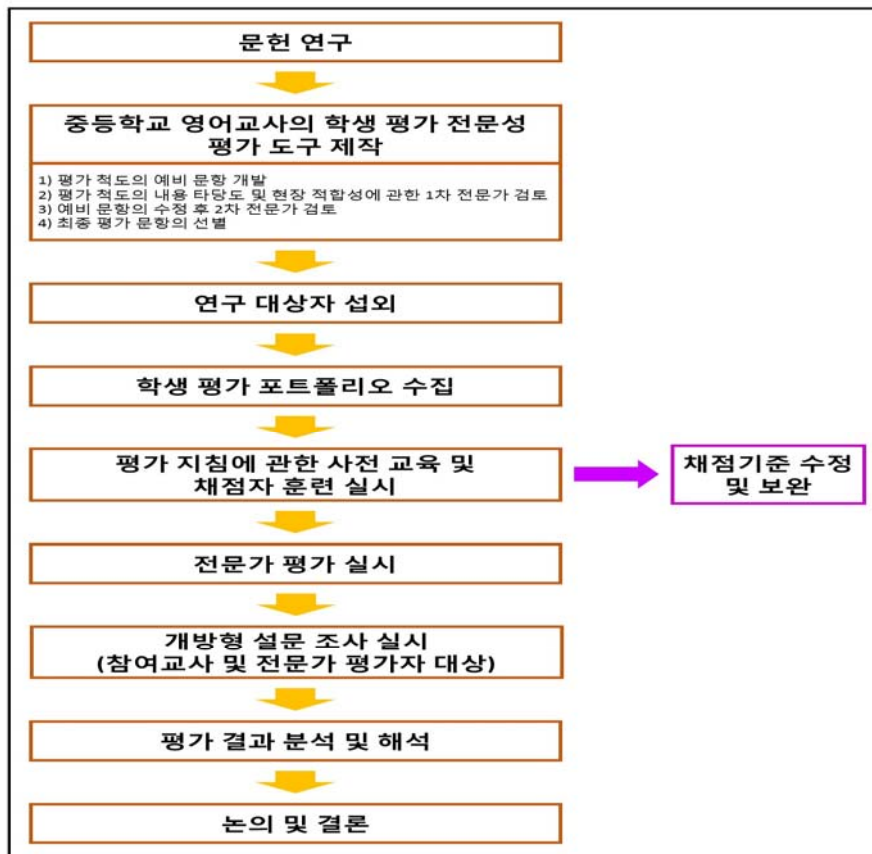
육 공동체 일원에 대한 고려가 부족하였다(주세형, 2011). 여섯째, 대부분의 연구에서 자기 보고식 설문조사를 활용한 자기 평가 방법을 활용하고 있어 측정의 신뢰도에 문제점이 발견되었다. 실제로, 교사들은 자신의 학생평가 전문성이 높다고 생각하는 경향이 있어(백순근, 2006; 이인제 외, 2004; 황은희, 백순근, 2008) 자기 평가 방법으로 교사의 학생평가 전문성에 관한 정확한 진단이 어렵다는 한계점을 지니고 있다. 마지막으로, 교사의 학생평가 전문성 측정의 타당도를 높이기 위해 설문 조사와 같은 단기간의 일회성 평가 보다는 장기간의 교사의 평가 활동 전반에 관한 실천적 측면을 평가할 필요가 있다.

따라서, 이 연구에서는 선행연구에서 드러난 여러 가지 한계점을 보완하기 위하여 중등학교 영어교사로 과목과 평가 대상을 한정하였으며, 평가 전문성의 인지적 요소 뿐만 아니라 보다 포괄적인 측면에서의 정의적 요소를 고려하여 실제적이고 구체적인 학생평가 전문성 평가도구를 개발하였다. 또한, 한 학기 동안 수집된 교사의 실제적인 평가 활동을 반영한 ‘학생평가 포트폴리오’를 바탕으로 ‘전문가 평가 방법’을 활용하여 평가의 타당도와 신뢰도를 높이하고자 하였다. 이 연구는 다른 교과목 또는 학교 급의 교사들을 위한 학생평가 전문성 평가도구 개발의 기초적인 자료를 제공할 것이며 교사의 평가 전문성 신장을 위한 교원 양성 과정이나 전·현직 교사교육에도 활용될 수 있을 것으로 기대된다.

Ⅲ. 연구 방법

1. 연구 절차

이 연구는 중등학교 영어교사의 실천적 학생평가 전문성을 평가 할 수 있는 도구를 개발하고 학생평가 포트폴리오를 활용한 전문가 평가 결과를 분석하여 개발된 평가도구의 타당도, 신뢰도 및 적합도를 검증하고자 하였다. 전반적인 연구 절차를 정리하면 [그림 Ⅲ-1]과 같다.



[그림 Ⅲ-1] 연구 절차

이 연구의 평가도구는 수집된 학생평가 포트폴리오를 평가 자료로 활용하여 중등학교 영어교사의 학생평가 전문성을 평가하고자 개발된 것이다. 교사의 학생평가 전문성 및 교사 포트폴리오와 관련된 국내·외의 다양한 문헌자료 및 선행연구 분석을 바탕으로 중등학교 영어교사의 학생평가 전문성 평가척도, 학생평가 포트폴리오의 구성요소 및 전문가 평정을 위한 채점기준을 개발하였다. 현직 교사 경력이 10년 이상이며 박사 수료 이상의 중등 영어교사 3인, 영어 교육학 전공 교수 1인과 시간강사 1인으로 구성된 영어교육 전문가 5인으로부터 평가척도의 하위 영역 및 예비 평가 문항, 학생평가 포트폴리오의 구성요소 및 세부 내용, 전문가 채점기준의 내용 타당도 및 현장 적합성에 관한 1차 전문가 검토 과정을 거쳤다. 1차 전문가 검토 결과를 반영하여 평가척도를 수정한 후 다시 2차 전문가 검토 과정을 거쳐 5개의 하위 영역으로 구성된 25개의 최종 평가 기준을 선정하였다.

평가척도를 개발한 다음 평가척도의 각 문항이 평가하고자 하는 내용을 해당 학생평가 포트폴리오 자료가 제대로 반영하고 있는지를 확인할 필요가 있다. 이를 위해 동일한 전문가 집단을 대상으로 한 전문가 협의회에서 평가 자료의 적절성에 관한 내용 타당도 검증 실시하였으며, 세 문항을 제외하고 대체적으로 평가 자료가 적합한 것으로 분석되었다. 구체적으로 전문가 집단이 평가 자료의 적절성이 비교적 낮다고 응답한 문항을 살펴보면 다음과 같다. 첫째, 학부모와의 의사소통 문항(“교사는 평가와 관련된 정보를 학부모와 공유하고 학생들의 학습 개선을 지원할 수 있는 방법을 제안하는가?”)의 경우, 평가 결과에 관한 학부모와의 의사소통이 주로 전화나 면담의 형식으로 이루어진다는 점에서 학교 현장에서 교사가 학부모와의 실제 의사소통 자료를 수집하기 어려운 점이 지적되었다. 둘째, ‘평가도구 개발’ 영역의 ‘평가도구 관리’ 문항

(“교사는 평가 문항의 특성과 평가도구의 양호도의 개념을 이해하고 확인하여 차후 평가도구 개발 자료로 활용 할 수 있는가?”)의 경우 차후 평가도구 개발 자료로 활용할 수 있는 역량을 해당 학기의 포트폴리오 자료만으로 파악하는데 한계점이 있다는 의견이 있었다. 셋째, ‘평가 결과 분석 및 활용’ 영역을 구성하는 ‘결과 분석 및 해석’ 항목(“교사는 개발한 평가도구의 양적·질적 평가 결과에 관한 기초적인 통계자료를 분석하고 그 의미를 정확하고 객관적으로 해석할 수 있는가?”)을 교과 협의록만을 활용하여 평가하기에는 어려움이 있다는 일부 전문가의 의견이 있었다. 같은 학교에서 같은 학년을 가르치는 교사의 경우 동일한 교과 협의록 자료를 제출한다는 점에서 해당 평가 자료만을 가지고 개별 교사들의 평가 결과 분석 및 해석 능력을 파악하기 어려운 점이 지적되었다. 평가 자료에서 파악하기 어려운 내용은 교사 설문지나 교사 포트폴리오에서 평가 저널의 내용을 참고하여 종합적으로 판단하였으며, 이러한 전문가 의견을 종합하여 평가도구를 최종적으로 수정·완성하였다.

다음으로, 수도권(서울, 경기, 인천) 지역의 중·고등학교에서 근무하고 있는 영어교사 중 연구 대상을 선정한 후, 한 학기(2016년도 2학기) 동안 학생평가 활동과 관련된 교사 포트폴리오를 수합하였다. 교사의 학생평가 포트폴리오에 관한 데이터 수집이 완료된 후, 3인의 전문가 평가자를 대상으로 사전교육 및 채점자 훈련을 실시하였다. 이 과정에서 나온 전문가 의견을 바탕으로 채점기준을 수정·보완한 후 중등학교 영어교사의 학생평가 전문성에 관한 전문가 평가를 실시하였다.

특히, 이 연구에서 학생평가 포트폴리오에 대한 전문가 평정 과정은 교사의 학생평가 전문성 평가의 전반적인 타당도와 신뢰도 확보에 있어서 매우 중요하므로 구체적인 절차를 살펴볼 필요가 있다. 먼저, 전문가 평가자는 교육 측정·평가 전공의 박사 학위 소지자 1인과 박사 수료

이상의 교직 경력을 지닌 영어 평가 전문가 2인을 포함한 총 3인으로 구성하였다. 다음으로, 전문가 평가자들을 대상으로 사전 교육 및 채점자 훈련을 실시하였다.

채점자 훈련의 목표는 크게 자기 일관성(self-consistency)을 위한 채점자 내 훈련과 평가가 궁극적으로 측정하려고 의도한 피험자의 능력에 관한 채점자들 간의 상호 일치성(agreement)을 유지, 발달시키는 것으로 나눌 수 있다. 채점자 훈련에서는 성취 수준 기술(PLD, Proficiency Level Descriptor)을 정확하고 일관되게 인지하는 것이 매우 중요하며(유진은, 2012), 공동 채점이 개별 채점보다 신뢰도가 높은 결과를 가져옴에 따라 채점자 간의 공동체 의식(sense of community)의 확립이 필요하다(이영식, 2000). 이를 바탕으로 이 연구에서는 전문가 평가자들에게 구체적이고 명확한 평가척도 및 채점기준(rubric)을 제시하려고 노력하였으며 개별 채점보다는 공동 채점을 원칙으로 평가 영역별로 교사별 차이를 분석하기 위한 분석적 채점 방식(analytic scoring)을 활용하였다.

구체적인 채점자 훈련 절차를 제시하면 다음과 같다. 먼저, 평가 목적 및 필요성, 평가척도의 구성요소 및 세부 항목, 채점기준에 관한 사전 교육 자료를 이메일을 통해 미리 배부하여 전문가 평가자들이 평가도구에 대해 충분히 숙지할 수 있도록 하였다. 2017년 3월 10일에 전문가 회의를 통해 평가도구에 대한 이해를 점검하였으며 각 전문가에게 수집된 학생평가 포트폴리오를 3개씩 배부하여 모의 채점 훈련을 실시하였다. 먼저, 전문가 평가자들은 학생평가 전문성의 하위 영역별로 5개의 세부 평가 문항에 대해 개별적으로 채점하고 논의 사항을 각자 기록하였다. 각 영역별 채점이 끝난 후, 자신의 평정 결과를 다른 전문가들과 비교·대조해 보는 과정을 거쳤다. 전문가 평가자들은 자신의 평정 이유를

채점기준을 바탕으로 설명하고 평정 결과가 일치하지 않는 문항은 따로 표시해 두었다. 최종적인 피드백 단계로서 모호한 채점기준에 관해 토론하고 평정 결과가 일치하지 않은 문항에 대해서는 재채점을 통해 평가 결과를 비교하는 등 채점자간 활발한 협의과정을 거쳤다. 채점자 훈련 과정에서 도출된 전문가 의견을 반영하여 2017년 3월 11일~12일 이틀간 36명의 중등학교 영어교사들의 학생평가 포트폴리오에 관한 전문가 평가가 실시되었다.

전문가 평가를 완료한 후 평가 기준 및 교사 배경 설문지(<부록 1> 참조)를 통해 수집된 교사 개인별, 학교별 특성에 따른 전문가 평가 결과를 구체적으로 분석하였다. 평가도구의 내용 타당도를 분석하기 위하여 영어교육 전문가 집단에게 각 문항이 학생평가 포트폴리오를 활용하여 학생평가 전문성의 각 구성요소를 제대로 측정하고 있는지에 대해 Likert형 5점 척도로 평정하도록 하였으며, 고전검사이론과 일반화가능도 이론을 각각 활용하여 개발한 평가도구의 신뢰도를 검증하였다. 또한, 평가도구의 적합도에 관해 분석하기 위해 연구 참여 교사 및 학생평가 포트폴리오의 전문가 평가자들을 대상으로 개방형 설문조사를 실시하였다. 먼저, 연구에 참여한 영어교사들을 대상으로 한 주요 설문 내용은 학생평가 포트폴리오의 장점 및 어려웠던 점, 학생평가 활동에 있어서 가장 중요하다고 생각하는 부분, 평가 관련 교사 연수에 관한 교사들의 의견 등에 관한 것이었다(<부록 2> 참조). 전문가 평가자를 대상으로는 학생평가 포트폴리오 평정 시 어려웠던 점, 평가도구에 관한 의견 및 학생평가 포트폴리오를 활용한 평가에서 보완해야 할 점 등을 묻는 서술형 설문조사를 실시하였다(<부록 3>참조). 최종적으로, 분석한 연구 결과를 바탕으로 연구의 논의 및 결론을 작성하였다.

2. 연구 대상

교사의 학생평가 전문성 평가도구 개발을 위한 구체적인 연구 대상은
①연구 참여 교사 ②평가도구의 내용 타당도 검증을 위한 전문가 집단
③ 학생평가 포트폴리오 평가를 위한 전문가 평정자로 구분할 수 있다.

가. 연구 참여 교사

개발된 평가도구를 실증적으로 검증하기 위해 서울, 경기, 인천 지역에서 근무하는 영어교사 42명으로부터 학생평가 포트폴리오를 수집하였다. 평가 포트폴리오 자료의 누락이 비교적 많은 불성실한 응답자 3명을 제외한 39명을 연구 대상으로 포함시켰다.

학생평가 포트폴리오의 전문가 평정을 위한 채점자 훈련과 평가척도 및 채점기준 점검을 위해 그 중 3명을 임의로 선정하여 모의 채점을 실시하였으며, 최종적으로 분석에 포함한 영어교사는 36명이었다. 연구에 참여한 교사들의 교사별, 학교별 특성에 따른 표집수를 살펴보면 <표 III-1>과 같다.

<표 III-1> 연구 참여 교사의 표집수

구분		모의 채점 대상		결과 분석 대상	
		표집수	비율 (%)	표집수	비율 (%)
성별	남교사	1	33	6	17
	여교사	2	67	30	83
교직경력	5년 이하	-	-	9	25
	6년 이상~10년 이하	3	100	11	31
	11년 이상~15년 이하	-	-	11	31
	16년 이상~20년 이하	-	-	2	6
	21년 이상	-	-	3	8
학력	학사졸업	2	67	22	61
	석사수료이상	1	33	14	39
전공	영어교육	2	67	32	89
	영문학	1	33	4	11
평균 학급당 학생수	20명 이하	-	-	1	3
	21명~25명	1	33	9	25
	26명~30명	-	-	10	28
	31명~35명	2	67	7	19
	36명 이상	-	-	9	25
학교유형	중학교	3	100	15	42
	일반계고	-	-	17	47
	특목고	-	-	4	11
혁신학교 여부	유	2	67	9	25
	무	1	33	27	75
전체		3	100	36	100

나. 평가도구의 내용 타당도 검증을 위한 전문가 집단

교사의 학생평가 전문성 평가도구는 평가척도의 예비 문항, 학생평가 포트폴리오 구성요소 및 전문가 채점기준의 내용 타당도에 관한 두 차례의 전문가 협의회를 통해 최종적으로 개발되었다.

개발한 평가도구의 내용 타당도 검증을 위한 전문가 집단은 영어 교육 분야의 박사 수료 이상의 현직 영어교사 3인, 대학교 시간 강사 1인 및 대학 교수 1인을 포함하여 총 5인으로 구성되었다. 영어 교육 전문가 집단에 관한 구체적인 정보는 <표 III-2>와 같다.

<표 III-2> 영어 교육 전문가들의 인적 사항

전문가	구분	학력	비고
전문가 A	대학교수	영어교육학 박사	대학교 영어 교육학 과목 강의(3년)
전문가 B	시간강사	영어교육학 박사	대학교 영어 쓰기 과목 강의(2년)
전문가 C	영어교사	영어교육학 박사	중·고등학교 교직 경험(16년)
전문가 D	영어교사	영어교육학 박사수료	중·고등학교 교직경험(12년)
전문가 E	영어교사	영어교육학 박사수료	중·고등학교 교직경험(11년)

다. 학생평가 포트폴리오 평정을 위한 전문가 평가자

이 연구에서는 교사의 학생평가 전문성 평가의 객관성을 높이기 위해 전문가 평가 방법을 활용하였다. 중등학교 영어교사들로부터 수집된 학생평가 포트폴리오를 평정하기 위한 전문가 평가단은 박사 수료 이상의 교육 측정 및 평가 분야의 전문가 1인과 영어 교육 분야의 평가 전문가 2인을 포함하여 총 3인으로 구성되었다. 전문가 평가자들에 관한 구체적인 정보는 <표 III-3>과 같다.

<표 III-3> 학생평가 포트폴리오 평가자들의 인적 사항

전문가	구분	학력	비고
전문가 F	교육 측정·평가 전문가	교육학(측정·평가 전공) 박사	•대학 교육평가 과목 강의(1년)
전문가 G	영어 평가 전문가	영어 교육학 (영어평가 전공) 박사 수료	•대학교 입학사정관 경험(2년) •고등학교 교직 경험(2년)
전문가 H	영어 평가 전문가	영어 교육학 (영어평가 전공) 박사 수료	•국가영어능력시험(NEAT)2,3급 출제 위원 및 1 급 채점위원 •중·고등학교 교 직경험(8년 이상)

3. 자료 분석 방법

가. 평가도구의 타당도 검증

영어 교육 전공 교수 1인, 시간강사 1인과 박사 수료 이상의 현직 영어교사 3인으로 구성된 영어 교육 전문가 집단을 대상으로 개발한 중등학교 영어교사의 학생평가 전문성 평가도구의 내용 타당도와 관련된 설문조사를 실시하였다. 구체적으로, 해당 학생평가 포트폴리오를 평가 자료로 활용하여 평가척도의 각 문항이 학생평가 전문성의 하위 구성 요인을 측정하는데 적합한지에 대한 전문가 의견을 이메일을 통해 물어보았다. 평가도구의 내용 타당도에 관한 설문조사는 Likert형 5점 척도(1: 전혀 적절하지 않다, 2: 적절하지 않다, 3: 보통이다, 4: 적절하다, 5: 매우 적절하다)로 응답하도록 구성하였으며, 2점 이하로 문항의 적절성이 낮다고 응답한 경우, 그 이유에 관해서 구체적으로 서술하도록 하였다. 5인의 전문가 집단이 평정한 점수를 평균하여 평가도구의 내용 타당도를 분석하였으며 응답 결과를 바탕으로 최종적으로 평가도구를 수정·보완하였다.

나. 평가도구의 신뢰도 검증

다음으로, 고전검사이론 및 일반화가능도이론을 각각 활용하여 평가도구의 신뢰도에 대한 검증을 실시하였다.

1) 고전검사이론을 활용한 신뢰도 검증

신뢰도는 검사도구가 측정하고자 하는 것을 얼마나 일관성 있게 측정하는지를 의미하는 것으로 크게 문항 내적 일관성(internal consistency)과 평가 시기나 평가자를 달리 했을 때 각 평가 결과의 일치 정도를 의미하는 안정성(stability)을 통해 확인할 수 있다. 고전 검사 이론에서 신뢰도는 진점수의 분산(σ_T^2)을 관찰점수의 분산(σ_X^2)으로 나눈 비율로 정의 할 수 있는데 이를 수식으로 나타내면 다음과 같다.

$$\rho^2 = \frac{\sigma_T^2}{\sigma_X^2} = 1 - \frac{\sigma_e^2}{\sigma_X^2}$$

신뢰도를 측정하는 방법으로 문항간의 일관성을 측정하는 크론바하 알파(Cronbach's α) 계수를 주로 많이 사용하며 이를 수식으로 나타내면 다음과 같다.

$$\alpha = \frac{k}{k-1} \left(1 - \frac{\sum_{j=1}^k \sigma_j^2}{\sigma_X^2} \right), \quad \sigma_j^2: j\text{번째 문항의 분산}$$

일반적으로 크론바하 알파(Cronbach's α) 계수가 .6이상일 경우 척도의 신뢰도가 비교적 양호한 것(acceptable)으로 판단 할 수 있다(이기훈, 2014). 이 연구에서는 교사의 학생평가 전문성의 각 하위 영역별

문항의 내적 신뢰도를 검증하기 위해 Cronbach's α 계수를 산출하였으며, 3인의 전문가 평가자 간의 채점자간 일치도(inter-rater reliability)를 구하기 위하여 전문가 평가자 2명씩 각각 짝을 이루어 Pearson의 단순적률상관계수를 활용하여 분석을 실시하였다.

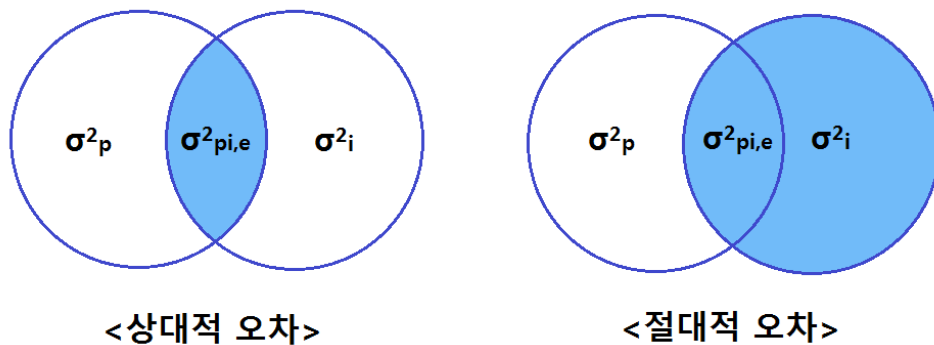
2) 일반화가능도이론을 활용한 신뢰도 검증

다음으로, 일반화가능도이론(Generalizability theory)을 통해 평가도구의 신뢰도를 검증하였다. 일반화가능도 이론은 Cronbach α 계수를 통한 문항 내적 신뢰도와 채점자 간 상관계수를 통해 채점자간 일치도를 분석한 기존의 고전검사 이론이 진점수를 고정시킴으로써 다양한 오차 요인에 대한 분석이 어렵다는 한계점을 극복한 것이다(양지승, 이규민, 2007). 즉, 일반화가능도이론은 측정 상황에서 발생하는 다양한 오차 요인 및 오차 분산의 상대적인 영향력을 산출할 뿐만 아니라 의사결정시 안정적인 점수를 얻기 위한 효율적인 측정 조건 또는 수준을 제시할 수 있다는 장점을 지닌다.

일반화가능도이론은 일반화 연구(G 연구)와 결정연구(D 연구)로 나뉘는데 먼저, G 연구는 허용 가능한 관찰전집(universe of admissible observation)을 추정하기 위해서 연구자가 설정한 설계에서 각 효과의 분산성분 추정값을 산출하여 다양한 오차요인의 상대적인 영향력을 비교하는 과정이다. 다음으로, D 연구는 G 연구 결과 허용 가능한 관찰전집의 모든 조건을 포함하거나 연구자의 관심에 따라 특정한 조건을 포함하는 일반화 전집(universe of generalization)을 정한다. 측정에 대한 해석이 상대적인지 절대적인지 명확히 하여 측정 오차를 정의한 후, 오차

추정의 종류에 따라 G 계수(generalizability coefficient)나 D 계수(dependability coefficient)를 산출한다. 이를 통해 적정 수준의 신뢰도를 확보하기 위한 효율적인 측정 국면의 조건을 제시하는 과정이다(김성숙, 김양분, 2001; 신동일, 2001; 이선영 외, 2015; Brennan, 2001).

오차 추정의 종류는 크게 상대적 결정과 절대적 결정으로 나뉘는데 상대적 결정은 다른 피험자와 비교하였을 때 개인의 상대적인 위치에 관심을 두는 것을 의미하는데 반해, 절대적 결정은 다른 피험자들과는 상관없이 개인의 지식, 기술, 태도 등에 대한 절대적인 수준을 매기는 것을 의미한다. [그림 III-2]에서 보듯이 주로 절대적 결정에 의한 오차가 상대적 결정에 의한 오차에 비해 더 많은 오차원의 수를 지니므로 절대적 결정에 의해 추정되는 D 계수가 상대적 오차에 따라 추정되는 G 계수에 비해 작은 경향이 있다(송일섭, 김수란, 2012).



[그림 III-2] 상대적 오차 및 절대적 오차 (송일섭, 김수란, 2012, p.115)

이 연구에서는 GENOVA(Brennan, 1983) 프로그램을 활용하여 일반 화가능력연구를 실시하였다. 세 명의 채점자(r)가 각각 동일한 36명의 영어교사들(p)의 학생평가 전문성을 평가 문항(i)을 통해 평가하므로 피

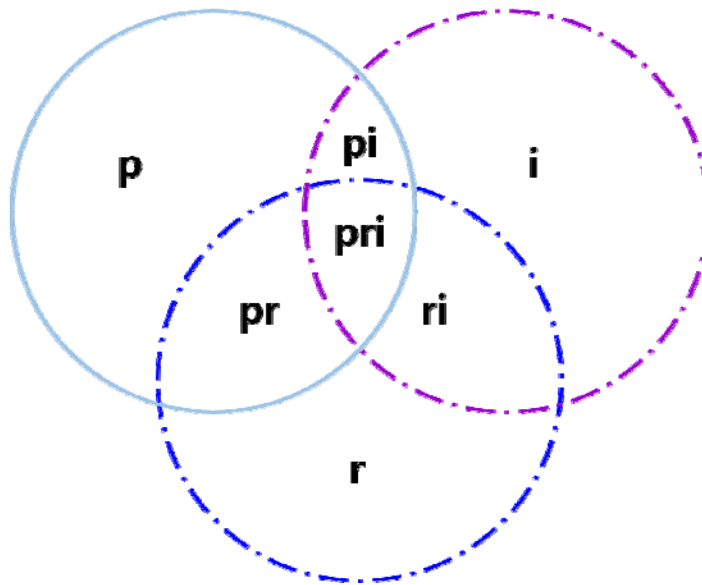
험자(P), 채점자(R) 및 평가 문항(I) 국면이 서로 교차된 $p \times i \times r$ 설계를 활용하여 G 연구를 실시하였다. 연구 모형에 대한 자료 배열 형태를 간략하게 나타내면 <표 III-4>와 같다.

<표 III-4> 연구 모형의 자료 배열

	채점자1					채점자2					채점자3				
	i1	i2	...	i24	i25	i1	i2	...	i24	i25	i1	i2	...	i24	i25
피험자1															
피험자2															
피험자3															
피험자4															
피험자5															
피험자6															
피험자7															
피험자8															
피험자9															
피험자10															
.															
.															
.															
.															
.															
피험자26															
피험자27															
피험자28															
피험자29															
피험자30															
피험자31															
피험자32															
피험자33															
피험자34															
피험자35															
피험자36															

G 연구에서는 산출된 피험자 국면(P), 채점자 국면(R), 평가 하위

문항 국면(I) 및 각 국면 간의 상호작용의 분산 성분 추정치를 산출하고 영어교사의 학생평가 전문성에 영향을 미치는 각 분산 성분 간의 상대적인 크기를 비교할 수 있다. 설계한 분석 모형을 벤다이어그램으로 나타내면 [그림 III-3]과 같다.



[그림 III-3] p x i x r 설계의 벤다이어그램

특정 영어교사의 학생평가 전문성 점수 ‘X’의 분산 성분을 수식으로 나타내면 다음과 같다.

$$\sigma_X^2 = \sigma_p^2 + \sigma_r^2 + \sigma_i^2 + \sigma_{pr}^2 + \sigma_{pi}^2 + \sigma_{ri}^2 + \sigma_{pri,e}^2$$

G 연구 결과를 바탕으로 한 D 연구에서는 p X I X R 설계를 바탕으로 상대오차 점수 분산을 고려한 G 계수를 산출할 수 있으며 이를 수식으로 나타내면 다음과 같다.

$$\rho^2 = \frac{\sigma_p^2}{\sigma_p^2 + \sigma_\delta^2}$$

(σ_p^2 : 전집점수분산, σ_δ^2 : 상대오차점수 분산)

중등학교 영어교사의 학생평가 전문성 평가도구의 경우 영어교사들 간 학생평가 전문성의 비교 뿐만 아니라 교사들의 학생평가 전문성이 기준 점수(cut-off score)에 도달하였는지 여부를 평가하는 것을 목적으로 하므로 준거 참조 검사의 성격을 또한 지닌다고 할 수 있다. 따라서, 준거 참조 검사의 신뢰도를 검증하기 위한 방법으로 절대오차 분산을 고려한 D 계수를 산출할 수 있으며 이를 수식으로 나타내면 다음과 같다.

$$\phi = \frac{\sigma_p^2}{\sigma_p^2 + \sigma_\Delta^2}$$

(σ_p^2 : 전집점수분산, σ_Δ^2 : 절대오차점수 분산)

일반적으로 G 계수가 0.8이상, D 계수의 경우 0.7 이상을 확보할 경우 적정 수준의 신뢰도로 판단할 수 있다(Brennan, 2001).

G 계수 및 D 계수를 산출하여 평가도구의 신뢰도를 검증한 후 채점자 및 평가 하위 문항의 수를 조절하여 서로 다른 조건에 따른 G 계수와 D 계수의 변화를 살펴본 후 시간과 비용의 측면을 고려할 때 적정한 수준의 신뢰도 계수를 확보할 수 있는 각 국면의 측정조건을 제시하였다. 이 연구에서는 채점자 수를 각각 2, 3, 4, 5, 6명으로, 평가 하위 문항의 수를 15, 20, 25, 30, 35, 40문항으로 조절하였을 때의 G 계수와 D 계수의 변화를 살펴보았다.

다. 평가도구의 적합도 검증

평가도구의 적합도(propersness)란 평가도구가 특정한 평가의 본질적인 목적이나 취지에 적합한 정도를 의미한다(백순근, 2001). 이 연구에서는 학생평가 포트폴리오를 활용한 평가가 중등학교 영어교사들의 실제적인 학생평가 전문성을 평가하는데 적합한지 여부를 검증하기 위하여 연구에 참여한 중등학교 영어교사와 학생평가 포트폴리오의 전문가 평정자를 대상으로 개방형 설문조사를 실시하였다(<부록 2>와 <부록 3> 참조).

연구 참여교사를 대상으로 한 설문조사의 주요 내용은 학생평가 포트폴리오 평가에서 가장 힘들었던 점, 데이터 수집상 어려웠던 항목, 학생평가 포트폴리오 평가를 통해 배우거나 느낀 점, 교사의 학생평가 활동에서 가장 중요하다고 생각하는 부분, 기존의 평가 관련 교사 연수에 관한 교사들의 의견에 관한 것이었다. 한편, 전문가 평정자를 대상으로 실시한 설문조사의 경우 학생평가 포트폴리오 평정 시 가장 어려웠던 점, 평가척도에서 불필요하거나 평정이 어려웠던 항목 및 해당 평가 도구에 보완되어야 할 점에 관한 질문을 포함하였다. 연구 참여 교사 및 전문가 평정자들의 응답 결과를 바탕으로 평가도구의 적합도를 긍정적인 측면과 부정적인 측면에서 다각도로 분석하고자 하였다.

IV. 연구 결과

이 연구는 학생평가 포트폴리오를 활용하여 중등학교 영어교사의 학생평가 전문성을 평가하기 위한 도구를 개발한 후, 수도권 소재 중·고등학교에서 근무하고 있는 영어교사 36명을 대상으로 실제 데이터를 수집하여 평가도구의 타당도, 신뢰도, 적합도를 분석함으로써 타당화 연구를 수행하였다.

먼저, 교사의 학생평가 전문성 평가도구는 다섯 가지 하위 영역으로 구성되었으며 각 평가 영역별로 5개의 세부 평가 기준을 포함하도록 개발하였다. 다음으로 평가도구의 타당화 연구의 일환으로 구체적인 평가 기준 및 교사 배경변수에 따른 교사의 학생평가 전문성에 대한 기술통계 분석 결과를 제시한 다음, 개발한 평가도구의 내용타당도와 신뢰도에 관한 검증을 실시하였다. 최종적으로 연구 참여 교사 및 전문가 평가자들을 대상으로 실시한 개방형 설문조사를 바탕으로 평가도구의 적합도에 관한 분석을 실시하였다.

1. 교사의 학생평가 전문성 평가도구의 개발

이 연구의 목적은 학생평가 포트폴리오를 활용하여 중등학교 영어교사의 실천적 학생평가 전문성을 평가하기 위한 도구를 개발하는 것이다. 이를 위해 ① ‘중등학교 영어교사의 학생평가 전문성 평가척도’, 평가자료로 활용되는 한 학기 동안 교사의 실제 학생평가 활동 전반에 관한 ② ‘학생평가 포트폴리오’의 구성요소, ③ 전문가 평정을 위한 ‘채점기준’을 각각 개발하였다.

가. 중등학교 영어교사의 학생평가 전문성 평가척도

‘중등학교 영어교사의 학생평가 전문성 평가척도’는 전문가 평가단이 한 학기 동안 수집된 학생평가 포트폴리오를 평가 자료로 활용하여 중등학교 영어교사들의 학생평가 전문성을 평정하기 위한 5점 척도의 체크리스트 형태로 개발되었다. 중등학교 영어교사의 학생평가 전문성 평가척도 개발 과정을 살펴보면 다음과 같다.

먼저, 중등학교 영어교사의 학생평가 전문성의 구성요인을 선정하고자 다양한 문헌조사와 선행연구를 분석하였다. 대부분의 선행연구에서 기준으로 삼고 있는 한국교육과정평가원의 ‘학생평가 전문성 기준’(김경희 외, 2006)의 구성요소(‘평가 방법의 선정’, ‘평가 도구의 개발’, ‘평가 실시·채점 및 성적 부여’, ‘평가 결과의 분석·해석·활용·의사소통’, ‘평가의 윤리성’)를 참고하여 학생평가 전문성의 5개의 하위 영역을 ‘평가 계획’, ‘평가도구 개발’, ‘평가 실행’, ‘평가 결과 분석 및 활용’, ‘평가에 대한 태도 및 인식’으로 재분류하였다.

구체적으로 살펴보면, 기존의 ‘평가방법의 선정’ 요인은 학습목표 및 평가목적에 적합한 평가방법을 선정할 수 있는 능력에 주로 중점을 둔 개념이었다. 이를 학습목표 및 평가목표 뿐만 아니라 교육과정, 학습자의 인지적, 정의적 특성에 관한 전반적인 이해, 실행 가능한 평가 활동에 관한 인식 등 본격적인 학생평가 활동에 앞서 교사가 이해하고 있거나 의사결정 할 수 있는 능력을 모두 포괄하여 ‘평가 계획’ 영역으로 개념을 확장하였다.

‘평가도구 개발’ 영역의 경우, 평가도구를 개발, 선택, 관리 할 수 있는 능력에 다양한 평가도구의 특징과 장·단점을 이해할 수 있는 능력 및 채점기준을 명확하게 설정하여 제시할 수 있는 능력을 함께 포함시켰

다.

다음으로 ‘평가 실시·채점 및 성적 부여’ 영역이 주로 채점과 성적 부여 과정에 중점을 둔 개념인 반면 평가 활동 수행 시 교사에게 요구되는 보다 넓은 의미의 학생평가 전문성을 ‘평가 실행’ 영역으로 포함시켰으며, ‘평가 결과의 분석·해석·활용·의사소통’ 영역은 ‘평가 결과 분석 및 활용’으로 명명하였다.

마지막으로, 기존의 ‘평가의 윤리성’ 항목이 주로 공정한 평가 활동을 위해 고려되는 교사의 학생 존중, 윤리적, 법적 책임에 초점을 두었다면 교사의 반성적 성찰 및 평가 개선 방안에 관한 모색, 평가 전문성 개발을 위한 노력, 동료교사와의 협업 능력, 평가의 환류적 기능에 대한 인식 등을 포함하여 교사의 학생평가 전문성의 정의적 요소를 다각도로 고려하여 ‘평가에 대한 태도 및 인식’ 영역을 구성하였다.

이와 같이 학생평가 전문성의 다섯 가지 구성요소를 선정한 후, 각 구성요소별 하위 평가 기준을 개발하기 위하여 ‘2009 개정 교육 과정에 따른 초·중학교 영어과 핵심 성취기준 개발 연구’ (임찬빈 외, 2013), ‘2009 개정 교육과정에 따른 고등학교 영어과 핵심 성취기준 개발 연구’ (임찬빈 외, 2014) 및 ‘2015 개정 교육과정에 따른 초·중학교 영어과 평가 기준 개발 연구’ (진경애 외, 2016)를 참조하였다. 그 외에도 교사의 학생평가 전문성과 영어 교과에서의 교사 전문성 또는 수업 평가와 관련된 국내·외 선행 연구들의 분석 결과를 바탕으로 세부적인 평가 항목을 개발하였다. 구체적인 내용을 살펴보면 다음과 같다.

첫째, 영어과 교육과정 중 특히, 2015 개정 교육 과정에서 영어과 평가도구의 개발 방향을 고려하여 평가척도의 세부 평가 기준을 설정하였다. 예를 들면, 2015 개정 교육과정에서 과정 평가에 유용한 서술형 평가, 수행평가 및 형성평가가 강조되면서 ‘평가 실행’ 영역에 교사가

성취 수준에 적합한 형성평가를 실시할 수 있는 능력을 평가 문항으로 포함하였다. 또한, 언어의 네 가지 기능별로 학생 수준을 고려한 다양한 평가도구를 개발할 수 있는 능력을 평가 기준으로 설정하였으며, 평가 방법이 다양해짐에 따라 다양한 형태의 학생 답안을 정확하고 일관성 있게 평가하기 위한 채점기준에 관한 평가 기준 역시 ‘평가도구 개발’ 영역에 포함하였다. 아울러, 학생들이 영어과 핵심 역량을 함양할 수 있도록 학생평가의 결과가 교수·학습 과정에 환류되어야 한다는 점이 강조되면서 ‘평가에 대한 태도 및 인식’ 영역에 ‘평가의 환류적 기능’에 관한 평가 기준을 개발하였다.

둘째, 교사의 학생평가 전문성과 관련된 대표적인 선행연구로는 교사의 학생평가 전문성 진단도구를 개발한 송미영, 김경희(2007)의 연구와 평가 포트폴리오를 활용하여 교사의 학생평가 전문성을 평가하기 위한 도구를 개발한 Webb의 연구(2012)를 포함하여 국내·외의 다양한 문헌연구를 참고하였다.

셋째, 영어 교과에서의 교사 전문성 및 수업 평가와 관련된 국내·외 선행 연구를 살펴보면 다음과 같다. 먼저, 대표적인 해외 교육 기관인 INTASC(The Interstate New Teacher Assessment and Support Consortium)의 초임 외국어 교사 기준¹⁾과 Cambridge ESOL의 TKT(Teaching Knowledge Test), TKT CLIL(Content and

1) *InTASC Model Core Teaching Standards(2011)*: ‘인간발달과 학습에 대한 지식’, ‘개별요구를 반영한 교수’, ‘동기부여와 교실운영기술’, ‘교과에 대한 지식’, ‘교과 내용의 적용’, ‘학생의 학습평가’, ‘교수 설계 기술’, ‘다양한 교수전략’, ‘전문적 수행과 윤리성’, ‘리더쉽과 협력’
 (http://www.google.co.kr/url?sa=t&rct=j&q=&esrc=s&frm=1&source=web&cd=1&ved=0ahUKEwj3-c20udvMAhXIUKYKHWi1B5gQFgggMAA&url=http%3A%2F%2Fwww.ccsso.org%2Fdocuments%2F2011%2Fintasc_model_core_teaching_standards_2011.pdf&usg=AFQjCNEanOK-juC1iM9DEGeDBAKMi9_sWA&sig2=5DnEQRUKNC7rK9b1b5dvfA&bvm=bv.122129774,d.dGY&cad=rjt)

Language Integrated Learning), CELTA(Certificate in English Language Teaching to Adults), CELTYL(Certificate in English Language Teaching to Young Learners), ICELT(International Certificate of English Language Teaching)의 바탕이 되는 영어교사 기준²⁾ 중 학생평가 관련 기준, 미국의 NCATE(The National Council for Accreditation of Teacher Education)에서 승인한 TESOL의 외국어(ESL) 영어교사 기준 중 평가 영역에 관한 루브릭³⁾ 및 영어교사 중심 평가 또는 교실 평가를 강조한 국외연구(Davison & Leung, 2009; Stanley & Murray, 2013)를 참고하였다.

다음으로, 국내 연구의 경우, 영어교사를 대상으로 교사 전문성을 분석한 다양한 선행 연구를 참고하였다. 구체적으로, 영어교사의 자질에 관한 연구(김은주, 2002; 이희경, 2007), 영어교사 교육, 교사 평가의 수행 기준 및 인식에 관한 연구(강양숙 외, 2015; 김혜영, 2007), 영어교사 전문성의 구성요소 및 전문성 발달을 고찰한 연구(김기택, 조규희, 2015; 백현정, 민찬규, 2009; 장경숙, 2008; 주순옥, 안경자, 2015; 하명애, 민찬규, 2008; Freeman, 1989; Richards, 2010; Shulman, 1987; Wallace, 1991), 초등 영어교사를 대상으로 교사 전문성의 정의적 영역을 평가하는 도구를 개발한 임수연의 연구(2014)등을 분석하였

2) Cambridge English Teaching Framework:

(<http://www.cambridgeenglish.org/teaching-english/cambridge-english-teaching-framework/>)

3) TESOL / NCATE Program Standards:

(

다. 영어교사의 전문성 중 특히, 학생평가 전문성에 초점을 둔 연구로는 영어교사의 학생평가 전문성의 구성요소를 고찰한 이인제 외(2004)의 연구와 ‘중등학교 영어과 교사의 학생평가 전문성 기준’을 바탕으로 한 연구(전병만 외, 2005; 전병만, 오준일, 2006)를 기초로 하였다. 이 외에도, 영어로 진행하는 영어 수업(TEE) 평가도구 개발과 관련된 연구(김진석, 2015; 진경애 외, 2009; 이병천, 김진석, 2011; 임찬빈 외, 2011; 최윤영 외, 2012) 및 영어과 수업 평가 기준(임찬빈 외, 2006)의 평가 관련 루브릭을 참고하였다.

우선, 앞서 살펴본 국내·외 문헌 연구 분석을 토대로 연구마다 공통적으로 언급하고 있는 항목을 바탕으로 다섯 가지 학생평가 전문성의 구성요소에 해당되는 항목을 모두 추려내어 65개의 1차 예비 문항을 개발하였다(<부록 4> 참조). 영어 교육 전문가 5인이 1차 전문가 협의회를 거쳐 예비 문항의 내용 타당도를 검증한 후 중복되는 내용을 삭제하고 각 구성요소별로 타당도가 높은 문항을 선별하여 예비 문항을 65개에서 43개 문항을 추려내었다. 중등학교 영어교사의 학생평가 전문성 평가척도의 2차 예비 문항은 <부록 5>에 첨부하였다. 43개의 예비 문항에 대한 2차 전문가 검토 및 논의 과정을 통해 중복되는 내용을 통합하거나 타당도와 현장 적합성이 낮은 문항을 삭제하여 최종적으로 25개의 평가 문항을 선별하였다.

다음으로, 평가도구의 내용 타당도를 검증하기 위하여 동일한 전문가 집단을 대상으로 이메일을 통해 각 문항이 해당 구성요소를 제대로 측정하는지에 대해 Likert형 5점 척도로 평정하도록 하였다. 전문가 집단의 의견을 토대로 모호하거나 부적절한 표현을 수정하여 최종적으로 5개 하위 영역의 25개의 세부 평가 기준을 지닌 평가척도로 완성하였다. 중등학교 영어교사의 학생평가 전문성 평가척도를 살펴보면, <표 IV-1>과 같다.

<표 IV-1> 중등학교 영어교사의 학생평가 전문성 평가척도

구성 요인	문항	평가 자료	전 혀 그 렇 지 않 다	그 렇 지 않 다	보 통 이 다	그 렇 다	매 우 그 렇 다
평 가 계 획	1.[교육과정] 교사는 해당 교육과정에서 제시하고 있는 영어과 성취 기준 및 평가 기준을 제대로 이해하고 있으며, 교육과정에 반영할 수 있는 역량을 지니고 있는가?	(A)(B) (C)(G) (L)	①	②	③	④	⑤
	2.[평가목표] 교사는 학습목표와 일치하는 평가목표를 분명히 설정하고 이를 구체적으로 제시하고 있는가?	(A)(D) (E)	①	②	③	④	⑤
	3.[평가의 실제성] 교사는 다양한 평가방법의 이론을 바탕으로 실행 가능한 평가 과업 및 평가 문항을 제시하고 있는가?	(A)(B) (C)(I)	①	②	③	④	⑤
	4.[학습자의 인지적특성] 교사는 영어 학습자의 인지적 특성 및 관련 학습 이론을 제대로 파악하고 있는가?	(A)(G) (H)(I) (L)(M)	①	②	③	④	⑤
	5.[학습자의 정의적특성] 교사는 영어 학습자의 다양한 정의적 특성을 이해하고 평가 계획에 반영하는가?	(A)(B) (G)(H) (I)(L) (N)	①	②	③	④	⑤
평 가 도 구 개 발	6.[평가도구이해]교사는 다양한 평가도구의 특징과 장·단점을 이해하고 있는가?	(A)(B) (C)(D) (G)(I) (L)(O)	①	②	③	④	⑤
	7.[평가도구개발]교사는 각 언어 기능별로 핵심 성취 기준의 달성 여부를 점검하는 다양한 평가도구를 직접 개발할 수 있는가?	(A)(D) (G)	①	②	③	④	⑤
	8.[평가도구선택]교사는 평가 목적에 따라 기존에 개발된 다양한 유형의 표준화 검사를 선택하고 활용할 수 있는가?	(A)(B) (C)(D) (E)	①	②	③	④	⑤
	9.[평가도구관리] 교사는 평가 문항의 특성과 평가도구의 타당도와 신뢰도의 개념을 이해하고 확인하여 차후 평가도구 개발 자료로 활용 할 수 있는가?	(A)(I)	①	②	③	④	⑤
	10.[채점기준] 교사는 개발된 평가 과업과 평가 문항의 채점기준을 명확하게 설정하여 제시하는가?	(A)(D)	①	②	③	④	⑤

구성 요인	문항	평가 자료	전 혀 그 렇 지 않 다	그 렇 지 않 다	보 통 이 다	그 렇 다	매 우 그 렇 다
평 가 실 행	11.[평가시기] 교사는 평가의 교육적 목적을 극대화 시키기 위해 적절한 시기에 평가를 실시하는가?	(A)(B) (C)(E) (G)(L)	①	②	③	④	⑤
	12.[학생안내] 교사는 영어학습자들에게 평가 실시 전에 미리 평가 시기, 평가도구, 채점기준, 평가 실시 방법 등에 대해 자세하게 안내하는가?	(E)	①	②	③	④	⑤
	13.[채점 및 성적부여] 교사는 학습 목표 및 평가 목표에 따라 적절한 채점 방식 및 점수 산출방식을 선택하여 실시하는가?	(A)(D) (G)(H) (I)(J) (K)(L)	①	②	③	④	⑤
	14.[숙제 및 형성평가] 교사는 수업 내용을 연습하거나 심화할 수 있는 과제를 제시하고 영어학습자들의 성취수준에 적합한 형성평가를 실시하는가?	(A)(G) (L)(M)	①	②	③	④	⑤
	15.[평가참여] 교사는 영어학습자들에게 자기평가 및 동료 간 상호 평가의 기회를 제공하는가?	(B)(G) (L)	①	②	③	④	⑤
평 가 결 과 분 석 및 활 용	16.[결과분석 및 해석] 교사는 개발한 평가도구의 양적·질적 평가 결과에 관한 기초적인 통계자료를 분석하고 그 의미를 정확하고 객관적으로 해석할 수 있는가?	(I)(O)	①	②	③	④	⑤
	17.[평가결과활용] 교사는 수집된 양적·질적 평가 결과를 활용하여 영어학습자의 학습 성취 정도에 대한 종합적 판단을 내릴 수 있으며 차기 평가 계획 및 평가 활동에 적용하는가?	(G)(I) (L)(M) (N)	①	②	③	④	⑤
	18.[학부모와의 의사소통] 교사는 평가 결과와 관련된 정보를 학부모와 공유하고 영어학습자들의 학습 개선을 지원할 수 있는 방법을 제안하는가?	(E)(M)	①	②	③	④	⑤
	19.[학생피드백] 교사는 평가 결과를 바탕으로 영어 학습에 대한 모니터링 및 영어학습자의 질문이나 숙제에 관한 피드백을 적절하게 제공하는가?	(G)(J) (K)(L) (M)(N)	①	②	③	④	⑤
	20.[교수기술 및 전략] 교사는 평가 결과를 바탕으로 수업의 내용, 수준(보충/심화), 교수 자료, 교수 방법 등을 조절하는가?	(I)(O)	①	②	③	④	⑤

구성 요인	문항	평가 자료	전 혀 그 렇 지 않 다	그 렇 지 않 다	보 통 이 다	그 렇 다	매 우 그 렇 다
평 가 에 대 한 태 도 및 인 식	21.[평가공정성]교사는 학생 존중을 바탕으로 학생 평가 활동 시에 학생 성적에 직·간접적으로 영향을 줄 수 있는 행위를 최소화하며 개별 영어학습자의 특성을 고려하여 평가도구, 학습 방법, 평가 기회 및 평가 방법을 공정하게 제공하는가?	(A)H (L)O	①	②	③	④	⑤
	22.[평가활동에 대한 반성]교사는 자신의 전반적인 교실내 평가 활동을 반성하고 발견된 문제점에 대한 대안을 모색하고 교수·학습 개선 방안을 스스로 제시 할 수 있는가?	(I)O	①	②	③	④	⑤
	23.[전문적 자율성]교사는 자신의 학생평가 전문성을 개발하기 위한 활동에 활발하게 참여하는가?	(F)	①	②	③	④	⑤
	24.[동료교사와의 협업능력] 교사는 평가 과정에서 교내·외 전문가 집단과 평가 절차, 채점기준, 평가 결과, 교수·학습 개선 방법 등에 대해 논의하고 협력하는가?	(D)(F) ①	①	②	③	④	⑤
	25.[평가의 환류적 기능] 교사의 전반적인 학생평가 활동이 영어학습자들의 핵심 역량 함양에 중점을 두고 긍정적 역류효과(positive washback)를 극대화시키는 방향으로 계획 및 실행되었는가?	전체 자료	①	②	③	④	⑤

중등학교 영어교사의 학생평가 전문성 평가척도는 교사별 학생평가 포트폴리오 자료를 활용하여 각 영역별로 5문항씩 총 25개의 문항으로 구성된 Likert형 5점 척도(① 전혀 그렇지 않다, ② 그렇지 않다, ③ 보통이다, ④ 그렇다, ⑤ 매우 그렇다)로 평정되었다. 해당 평가척도는 세부 평가 항목 개발 시 영어과 교육과정을 비롯하여 영어교과에서의 교사 전문성 및 수업 평가와 관련된 다양한 문헌 연구를 참고하였고 평가척도의 내용 타당도 검증을 영어 교육 전문가들을

대상으로 실시하였으며, 개발한 평가척도를 적용하여 중등학교 영어 교사들의 실제 데이터만을 바탕으로 교사의 학생평가 전문성을 분석하였다. 또한, 평가 자료로 활용되는 학생평가 포트폴리오의 구성요소에서도 영어교과와 특성을 고려하여 세부 내용을 명시하였으며, 전문가 평정을 위한 채점기준에서 각 평가 문항의 세부 채점기준의 적용 시 영어교과와 학생평가 특성이 많이 반영된다는 점에서 해당 평가척도는 영역 특수성을 지닌다고 할 수 있다.

한편, 교사의 학생평가 전문성 평가척도는 일부 평가 기준을 제외하고 영역 일반적인 내용을 또한 포함하고 있다. 따라서, 평가척도, 학생평가 포트폴리오 자료 및 채점기준을 해당 교과목 특성에 따라 수정한다면 다른 교과목 교사들의 학생평가 전문성 평가에도 활용할 수 있을 것이다.

나. 학생평가 포트폴리오 구성요소

교사의 학생평가 전문성에 대한 평가의 타당도와 신뢰도를 높이기 위해서는 설문 조사 또는 수업 관찰과 같은 단기간의 일회성 평가보다는 장기간의 교사의 실천적 학생평가 활동 전반에 관한 평가가 필요할 것이다. 따라서, 이 연구에서는 평가 계획에서부터 평가 실행 과정, 평가 실시 후 분석 및 활용에 이르기까지 교사의 전반적인 학생평가 활동에 관한 자료를 ‘학생평가 포트폴리오’에 포함시켜 교사의 실제적인(authentic) 학생평가 전문성을 평가하고자 하였다.

학생평가 포트폴리오를 수집하기에 앞서 교사 모집 단계에서 연구 참

여 교사들에게 연구 목적 및 학생평가 포트폴리오의 구성요소에 관한 충분한 이해를 돕기 위해 서면 또는 이메일을 통해 안내 자료를 배부하였다. 또한, 연구 참여 교사들의 심리적 부담감을 경감시켜주고 효율적인 자료 수집을 위해 학생평가 포트폴리오의 샘플을 직접 제작하여 각 학교에 배치할 수 있도록 하였다. 학생평가 포트폴리오는 2016년도 2학기(9월-12월)동안 교사별로 수집·제작되어 2017년 1월 초에 수거를 완료하였으며 이를 활용하여 ‘중등학교 영어교사의 학생평가 전문성 평가 척도’를 바탕으로 3인으로 구성된 전문가 평가 집단이 개별 교사들의 학생평가 전문성을 평정하였다.

학생평가 포트폴리오는 교사 포트폴리오 중에서 특히 교사의 실제적인 학생평가 활동에 관한 정보의 모음집으로 일정 기간 동안의 다양한 교수·학습 상황에서 교사의 성취 정도를 보여주는 학생평가 활동을 체계적으로 수집한 것이라고 할 수 있다(최금진, 2012). 학생평가 포트폴리오의 구성요소 및 세부 내용은 티칭 포트폴리오(teaching portfolio)의 구성요소를 참고하여 개발하였다. 티칭 포트폴리오의 구성요소는 크게 교수 활동의 성과와 노력을 보여주는 각종 증거 자료 및 자기 성찰 자료로 구분할 수 있다. 수업 활동을 문서화한 증거 자료는 다시 ①자신으로부터 수집되는 자료 ②타인으로부터 수집되는 자료 ③교수·학습의 산출물로 나눌 수 있다. 첫째, 자신으로부터 수집되는 자료는 수업계획서, 수업 기술 향상을 위한 프로그램 참여, 자신의 수업을 평가하고 개선하기 위한 노력들, 평가 방법과 평가도구 등 교사 자신이 수업을 향상시키고 평가하기 위한 단계와 과정을 기술하는 구성요소들로 구성된다. 둘째, 타인으로부터 수집되는 자료는 학생들로부터 수집된 수업자료, 강의 평가 결과, 동료교사나 학생으로부터 인정받은 사항, 교과 개발이나 수업 개선을 위해 동료교사에게 도움을 준 증거 등을 포함한다. 셋째,

교수·학습 산출물에는 채점 사례, 학생들의 과제나 학습 향상에 관한 코멘트, 학생 피드백 자료, 학생들의 학습 수행정도를 비교한 자료 등이 포함될 수 있다(조용개, 2009).

이 연구에서는 교사 포트폴리오의 구성요소와 관련된 선행연구(손영민, 2007; 조용개, 2009; 최금진, 2012)를 바탕으로 학생평가 포트폴리오의 구성요소를 ‘자신으로부터 수집되는 자료’, ‘타인으로부터 수집되는 자료’, ‘교수·학습 산출물’ 및 ‘자기 성찰 자료’의 네 가지 영역으로 구분하여 제시하였으며, 영어 교과의 특성을 고려하여 세부 내용을 명시하였다. 학생평가 포트폴리오의 구체적인 구성요소 및 내용을 살펴보면 <표 IV-2>와 같다.

‘과제 및 학습 태도 체크리스트’, ‘비공식적 학생평가 자료 및 기록물’, ‘학생평가 전문성 개발 활동 참여 기록’, ‘교과 협의록’ 및 ‘평가 저널’ 등 수집된 학생평가 포트폴리오의 실제 예시자료는 <부록 6>에 첨부하였다. 특히, 저널쓰기는 교사 스스로 자신의 문제점을 인식하고 대안을 생각해보는 기회를 가짐으로써 반성적 사고를 발달시키는데 효과적이다(석은조, 2006; 손영민, 2007). 이러한 점에서 교사의 반성적 사고 능력을 포함하여 학생평가 활동에 대한 전반적인 교사의 태도 및 인식 정도를 살펴보고자 학생평가 포트폴리오에 ‘평가 저널’을 포함시켰다.

〈표 IV-2〉 학생평가 포트폴리오의 구성요소

구분	자료명	평가 자료	세부 내용
자신 으로 부터 수집 되는 자료	㉠	해당 년도 영어 수업 계획서 및 평가계획서	영어과 핵심 성취 기준에 따른 2016년도 2학기 수행평가 및 지필평가의 평가 기준, 목표, 성취 수준, 평가 방법, 범위, 평가도구 개발, 채점방식, 성적 처리 등
	㉡	해당 학기 영어 수행평가 문제지	
	㉢	해당 학기 영어 중간 또는 기말고사 지필평가 문제지	
	㉣	영어 서술형 평가 채점기준 및 동일답안 인정 기준	초안, 교사 협의회 이후 수정안 모두 제출 ※ 채점기준표를 따로 작성하지 않을 경우 해당 학기 중간고사 또는 기말고사의 이원 목적 분류표로 대체
	㉤	수행 평가 및 지필 평가에 대한 학생 안내자료	평가 범위, 평가 기준 등에 관련된 학교 홈페이지 게시 내용 또는 학급 안내지
	㉥	해당 년도의 학생평가 전문성 개발 활동 참여 기록	2016년도에 참여한 평가 관련 연수(서술형 평가 채점 연수 포함), 워크샵, 교사 커뮤니티, 세미나, 스터디 그룹의 다양한 강의 자료, 평가 자료, 프로그램 브로셔, 수료증, 이수증 등
타인 으로 부터 수집 되는 자료	㉦	비공식적 학생평가 자료 및 기록물	단원평가지, 퀴즈 문제, 보충·심화 학습지 및 답안 샘플, 관찰기록지 등 평소의 학생평가 활동에 대한 정보
	㉧	학교의 성적 관리 규정 또는 영어과 성적 처리 규정	
	㉨	교과 협의록 및 평가 보고서 샘플	평가 계획 수립, 평가 결과 분석에 관한 영어교사 협의록 또는 교육청 보고용 평가 결과 보고서
교수·학 습 산 출 물	㉩	해당 학기 영어 수행 평가의 채점된 학생 답안 샘플	
	㉪	해당 학기 영어 중간고사 또는 기말고사 서술형 문항의 채점된 학생 답안 샘플	
	㉫	과제 및 학습태도 체크리스트	과제 점검 체크리스트, 과제에 관한 교사의 코멘트
	㉬	평가 후 교사의 피드백	쓰기 수행평가에 대한 첨삭, 학습지, 교과서에 기록한 피드백(단어 퀴즈, 숙제, 개별 필기활동, 집단 필기활동에 대한 교사의 표시, 도장, 채점, 기호 사용, 의견 제시 등), 성적통지표에서의 교사의 코멘트, 공식적·비공식적 가정 통신물, 모의고사나 성취도 평가 실시 후 교사의 피드백(<예시>오답노트작성, 개인별 코멘트, 이메일 등)
	㉭	학생생활기록부의 교과별 세부영역 기입 내용	영어학습자들의 학습과정 및 결과에 관한 영어교사의 피드백 및 학생생활기록부 기록
자기 성찰 자료	㉮	자기 평가 보고서(평가 저널)	한 학기 동안의 자신의 학생평가 활동에 관한 반성 및 느낀 점에 대해 구체적으로 서술(교육 철학, 평가 방법, 평가 목표, 잘한 점, 부족한 점, 자신의 평가 활동이 학생들의 학습 방법에 미치는 영향, 공정한 평가를 위한 노력 등)

다. 전문가 평정을 위한 채점기준

평가척도 개발 과정에서 참고한 다양한 문헌 연구 분석 및 전문가 협의과정을 거쳐 전문가 평가자를 위한 채점기준을 개발하였다. 학생평가 포트폴리오의 전문가 평가자는 각 영어교사에게 총 25개의 하위 평가 기준별 세부 채점기준에 따라 1점에서 5점 사이의 점수(5점 척도)를 부여하였다. 예를 들면, ‘평가 실행’ 영역의 12번 평가 문항인 ‘학생 안내’ 항목의 채점기준을 살펴보면 다음 <표 IV-3>과 같다.

<표 IV-3> ‘학생 안내’ 항목에 관한 세부 채점기준

세부 기준 점수	①평가 계획(시기 및 시간), ②평가도구(평가 과업, 과제 등), ③채점기준, ④평가 실시 방법, ⑤범위 등을 학생들에게 안내함.
1	미리 안내하지 않음.
2	1개 항목만 안내함.
3	2개 항목만 안내함.
4	1~2개 항목을 제외하고 안내함.
5	모두, 구체적으로 안내함.

‘학생 안내’ 기준, 즉 “교사는 학생들에게 평가 실시 전에 미리 평가 시기, 평가도구, 채점기준, 평가 실시 방법 등에 대해 자세하게 안내하는가?”에 대한 교사의 학생평가 전문성은 학생평가 포트폴리오 중 “수행평가 및 지필평가에 대한 학생 안내자료”를 토대로 <표 IV-3>의 채점기준에 의해 3명의 전문가가 평정한 점수를 평균한 값으로 측정되었다. 전체 평가 문항에 대한 구체적인 채점기준은 <부록 7>에 제시

하였다.

2. 교사의 학생평가 전문성 평가도구의 타당화

평가도구의 타당화는 개발한 평가도구를 활용하여 수집된 교사의 학생평가 포트폴리오의 전문가 평가 결과에 대한 기술 통계 분석, 설문조사 응답 분석을 통한 평가 도구의 타당도, 신뢰도 및 적합도 검증 과정을 통해 수행되었다.

2.1. 기술통계 분석 결과

먼저, 중등학교 영어교사의 학생평가 전문성의 구체적인 특성을 살펴보기 위해 평가 기준 및 교사 배경 변수에 따라 학생평가 포트폴리오를 활용한 전문가 평가를 실시한 결과를 분석하여 제시하였다.

가. 학생평가 전문성의 평가 기준별 기술통계 결과

이 연구에서는 중등학교 영어교사의 학생평가 전문성의 평가 기준을 크게 ‘평가계획’, ‘평가도구 개발’, ‘평가 실행’, ‘평가 결과 분석 및 활용’, ‘평가에 대한 태도 및 인식’의 5개의 하위 영역으로 구성하였으며, 각 하위 영역별로 5개씩의 세부적인 평가 기준을 활용하여 교사의

학생평가 전문성을 평가하였다. 교사의 학생평가 전문성의 총점 및 각 하위 영역별 기술통계 결과를 살펴보면 <표 IV-4>와 같다.

<표 IV-4> 학생평가 전문성 총점 및 하위 영역별 기술통계 결과

구분	최소값	최대값	평균	표준편차	왜도 (s.e)	첨도 (s.e)
평가계획	10.33	24.33	16.62	3.20	.29 (.39)	-.23 (.77)
평가도구개발	9.33	23.00	17.07	3.21	-.77 (.39)	.48 (.77)
평가실행	7.00	21.00	14.28	3.74	-.13 (.39)	-.90 (.77)
평가 결과 분석 및 활용	7.33	20.67	13.27	3.30	-.10 (.39)	-.53 (.77)
평가에 대한 태도 및 인식	7.67	22.67	13.19	3.71	.80 (.39)	.25 (.77)
전체총점	52.00	101.33	74.43	13.31	-.06 (.39)	-.94 (.77)

분석 결과, 교사의 학생평가 전문성의 전체 총점의 평균은 125점 만점에 74.43점이었으며 표준편차는 13.31, 최소값은 52.00점, 최대값은 101.33점으로 교사별로 큰 차이가 나타남을 알 수 있었다. 전반적으로 ‘평가 계획’ 과 ‘평가도구 개발’ 영역의 평균 수준이 높은 편이었으며 ‘평가실행’, ‘평가 결과 분석 및 활용’ 과 ‘평가에 대한 태도 및 인식’ 영역은 낮은 편이었다. 평균 수준이 낮은 세 영역 중 특히 ‘평가 실행’ 과 ‘평가에 대한 태도 및 인식’ 의 경우 표준편차가 각각 3.74와 3.71로 다른 영역에 비해 상대적으로 크게 나타났다. 이는 교사 간 학생평가 전문성의 차이가 해당 영역에서 더욱 크게 나타남을 의미한다. 교사의 학생평가 전문성의 총점 및 각 하위 영역별 점수 분포를 히스토그램으로 나타내면 [그림 IV-1] ~ [그림 IV-6]과 같다(<부록 8>첨부).

다음으로 학생평가 전문성의 세부 평가 기준별 기술통계 결과를 살펴보면 <표 IV-5>와 같다.

〈표 IV-5〉 학생평가 전문성의 세부 평가 기준별 기술통계 결과

구성요소	평가 기준	최소값	최대값	평균	표준 편차	왜도 (s.e)	첨도 (s.e)
평가계획	교육과정	1.00	5.00	3.30	1.20	-.19 (.39)	-1.04 (.77)
	평가목표	1.00	5.00	3.33	1.10	-.15 (.39)	-.41 (.77)
	평가의 실제성	4.33	5.00	4.96	0.13	-3.87 (.39)	15.55 (.77)
	학습자의 인지적특성	1.00	4.67	2.40	1.02	.52 (.39)	-.15 (.77)
	학습자의 정의적 특성	1.00	5.00	2.63	1.04	.49 (.39)	-.07 (.77)
평가도구개발	평가도구이해	1.00	5.00	3.43	1.00	-.60 (.39)	-.45 (.77)
	평가도구개발	2.00	5.00	3.73	0.83	-.75 (.39)	-.28 (.77)
	평가도구선택	1.67	5.00	3.69	0.94	.05 (.39)	-1.16 (.77)
	평가도구관리	1.00	4.67	2.22	1.09	.57 (.39)	-.74 (.77)
	채점기준	1.00	5.00	4.01	1.21	-1.22 (.39)	.44 (.77)
평가실행	평가시기	1.00	5.00	3.45	1.43	-.29 (.39)	-1.70 (.77)
	학생 안내	1.00	5.00	3.56	1.52	-.90 (.39)	-.86 (.77)
	채점 및 성적부여	1.00	5.00	2.74	1.18	.55 (.39)	-.72 (.77)
	숙제 및 형성평가	1.00	5.00	3.34	1.30	-.38 (.39)	-.90 (.77)
	평가참여	1.00	3.00	1.19	0.51	2.88 (.39)	7.31 (.77)
평가결과 분석 및 활용	평가결과 분석 및 해석	1.00	5.00	3.16	1.63	-.27 (.39)	-1.71 (.77)
	평가결과 활용	1.00	4.67	2.93	1.01	.03 (.39)	-1.16 (.77)
	학부모와의 의사소통	1.00	3.00	1.19	0.51	3.09 (.39)	8.80 (.77)
	학생 피드백	1.00	4.00	2.72	0.89	-.65 (.39)	-.43 (.77)
	교수기술 및 전략	1.00	5.00	3.28	0.86	-.47 (.39)	.16 (.77)

평가에 대한 태도 및 인식	평가공정성	1.00	5.00	3.00	1.15	.33 (.39)	-1.04 (.77)
	평가 활동에 대한 반성	1.00	5.00	3.09	0.91	-.35 (.39)	.15 (.77)
	전문적 자율성	1.00	5.00	1.74	1.19	1.81 (.39)	2.36 (.77)
	동료 교사와의 협업 능력	1.00	4.67	2.10	0.97	1.04 (.39)	.35 (.77)
	평가의 환류적 기능	1.67	5.00	3.25	0.84	.14 (.39)	-.77 (.77)

분석 결과, 교사들은 실행 가능한 평가 과업 및 평가 문항을 제시할 수 있는 능력인 ‘평가의 실체성’ 과 개발한 평가도구의 채점기준을 명확하게 설정하여 제시할 수 있는 능력인 ‘채점기준’ 항목에서 평균 수준이 높았다. 반면에, 학생들에게 자기평가 및 동료 간 상호 평가의 기회를 제공하는 ‘평가 참여’ 와 평가 결과를 학부모와 공유할 수 있는 능력인 ‘학부모와의 의사소통’ 기준의 평균 수준이 가장 낮은 것으로 나타났다. 이는 학생평가 전문성 기준 중 ‘학부모와의 의사소통’ 항목의 점수가 가장 낮게 나온 Plake et al.(1993)의 연구와 교사들의 학생평가 특성에 대한 높은 이해수준에 비해 ‘자기 평가의 기회 제공’ 과 ‘평가 결과에 대한 의사소통’ 영역에서는 실천적 학생평가 전문성이 상대적으로 낮게 나타난 김신영(2014)의 연구와 동일한 결과를 보여주었다.

또한, 세부 평가 기준별 평균 점수의 표준편차를 살펴보면 ‘평가 결과 분석 및 해석’ 항목의 표준편차가 1.63으로 가장 크게 나타났으며 ‘학생 안내’ 와 ‘평가시기’ 항목이 그 뒤를 따랐다. 따라서, 교사별로 평가 결과에 대해 분석하고 해석할 수 있는 능력에서 가장 큰 차이가 나타남을 알 수 있다.

나. 교사 배경변수에 따른 학생평가 전문성의 기술통계 결과

이 연구에서는 교사의 학생평가 전문성이 교사의 개인적 특성 및 근무하는 학교 특성에 따라 차이를 보일 것이라는 가정하에 중등학교 영어 교사의 학생평가 전문성 수준을 측정할 수 있는 평가도구를 개발하였다. 교사 배경 설문조사를 통해 실제 수집한 자료를 활용하여 교사의 성별, 교직 경력, 학력, 학교급을 중심으로 교사의 학생평가 전문성의 총점 및 각 하위 영역별 기술통계 결과를 제시하였다.

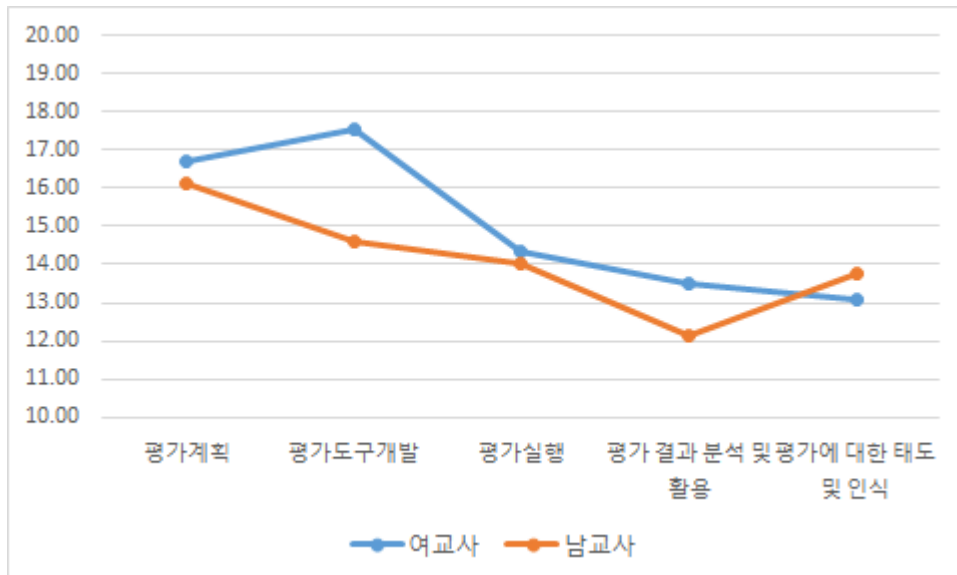
1) 성별에 따른 학생평가 전문성 분석 결과

연구에 참여한 총 36명의 영어교사들은 여교사가 30명, 남교사 6명으로 구성되어 있다. 먼저, 교사 개인적 특성 중 성별에 따른 학생평가 전문성의 기술통계 결과를 살펴보면 <표 IV-6>과 같다.

<표 IV-6> 성별에 따른 학생평가 전문성의 기술통계 결과

구성요소	여교사(N=30)		남교사(N=6)	
	평균	표준편차	평균	표준편차
평가계획	16.72	3.29	16.11	2.94
평가도구개발	17.57	2.83	14.61	4.10
평가 실행	14.33	3.72	14.00	4.18
평가결과 분석 및 활용	13.50	3.21	12.11	3.85
평가에 대한 태도 및 인식	13.07	3.73	13.78	3.85
총점	75.19	12.91	70.61	15.85

기술통계 분석 결과, ‘평가에 대한 태도 및 인식’ 영역의 경우 남교사 집단이 여교사 집단에 비해 평균점수가 더 높게 나타난 반면, 다른 4개의 하위 영역에서는 모두 여교사 집단의 학생평가 전문성이 상대적으로 더 높은 것으로 나타났다. 여교사 집단 내에서도 ‘평가에 대한 태도 및 인식’ 영역 과 ‘평가 실행’ 영역의 표준편차가 상대적으로 높은 편으로 해당 영역에서 학생평가 전문성의 개인차가 더욱 크게 나타남을 알 수 있다. 여교사와 남교사 간의 학생평가 전문성의 각 하위 영역별 평균 점수를 비교하면 [그림 IV-7]과 같다.



[그림 IV-7] 성별에 따른 교사의 학생평가 전문성 평균점수 비교

2) 교직경력에 따른 학생평가 전문성 분석 결과

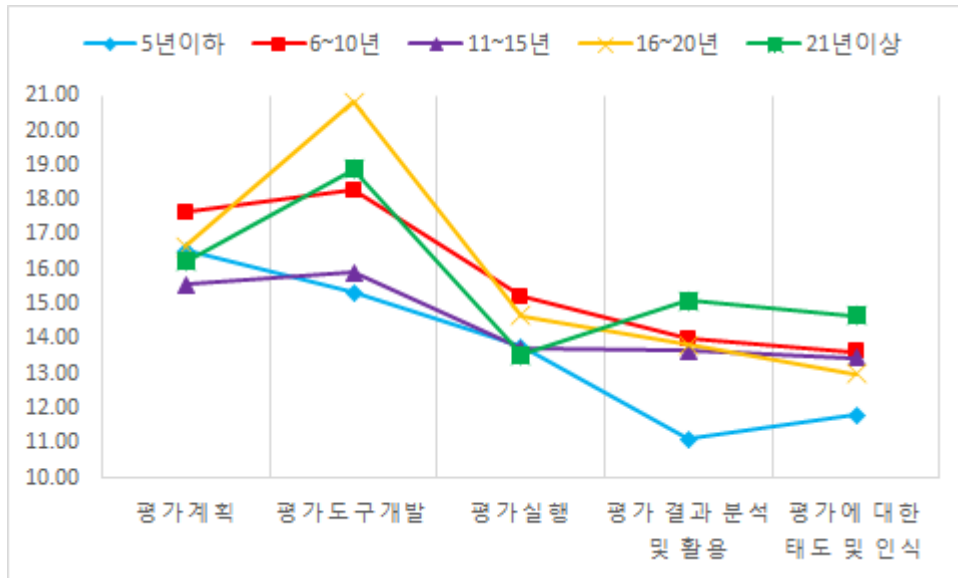
연구에 참여한 총 36명의 영어교사들을 교직경력에 따라 분류하면, 5년 이하의 경력을 지닌 교사 9명, 6년~10년 경력의 교사 12명, 11년~15년 경력의 교사 10명, 16년~20년 경력의 교사 2명, 21년 이상의 교직경력을 지닌 교사가 3명이었다. 교직경력에 따른 교사의 학생평가 전문성에 대한 기술통계 분석 결과를 살펴보면 <표 IV-7>과 같다.

<표 IV-7> 교직경력에 따른 학생평가 전문성의 기술통계 결과

구성 요소	5년 이하 (N=9)	6년~10년 (N=12)	11년~15년 (N=10)	16년~20년 (N=2)	21년 이상 (N=3)
	평균 (표준편차)	평균 (표준편차)	평균 (표준편차)	평균 (표준편차)	평균 (표준편차)
평가계획	16.56 (3.21)	17.64 (3.06)	15.57 (3.30)	16.67 (5.66)	16.22 (3.10)
평가 도구개발	15.33 (3.45)	18.28 (2.24)	15.90 (3.41)	20.83 (1.65)	18.89 (1.71)
평가 실행	13.78 (3.83)	15.22 (3.73)	13.73 (4.16)	14.67 (2.36)	13.56 (4.44)
평가결과 분석 및 활용	11.11 (2.95)	14.00 (3.36)	13.67 (3.33)	13.83 (3.06)	15.11 (2.99)
평가에 대한 태도 및 인식	11.81 (3.08)	13.64 (3.96)	13.47 (4.01)	13.00 (0.47)	14.67 (5.51)
총점	68.59 (13.02)	78.78 (13.16)	72.33 (14.33)	79.00 (3.77)	78.44 (14.25)

기술통계 분석 결과, ‘평가 계획’ 과 ‘평가 실행’ 영역에서는 6년~10년의 교직 경력을 지닌 교사 집단의 평균 점수가 가장 높게 나타났다. ‘평가도구 개발’ 영역의 경우, 16년~20년의 교직 경력을 지닌 집단에서 ‘평가 결과 분석 및 활용’ 과 ‘평가에 대한 태도 및 인식’ 영역에서는 21년 이상의 교직 경험이 풍부한 집단에서 상대적으로 가장 높은 점수를 보여주었다. 한편, 21년 이상의 교직 경험이 가장 풍부한 집단의 경우 실제 평가 활동을 수행하는 능력인 ‘평가 실행’ 영역의 점

수는 가장 낮은 것으로 나타났다. 또한, 5년 이하의 교직경험이 비교적 부족한 신규 교사들의 경우 대부분의 영역에서 가장 낮은 평균 점수를 나타냈다. 교직 경력에 따른 학생평가 전문성의 각 하위 영역별 평균 점수를 비교하면 [그림 IV-8]과 같다.



[그림 IV-8] 교직경력에 따른 교사의 학생평가 전문성 평균점수 비교

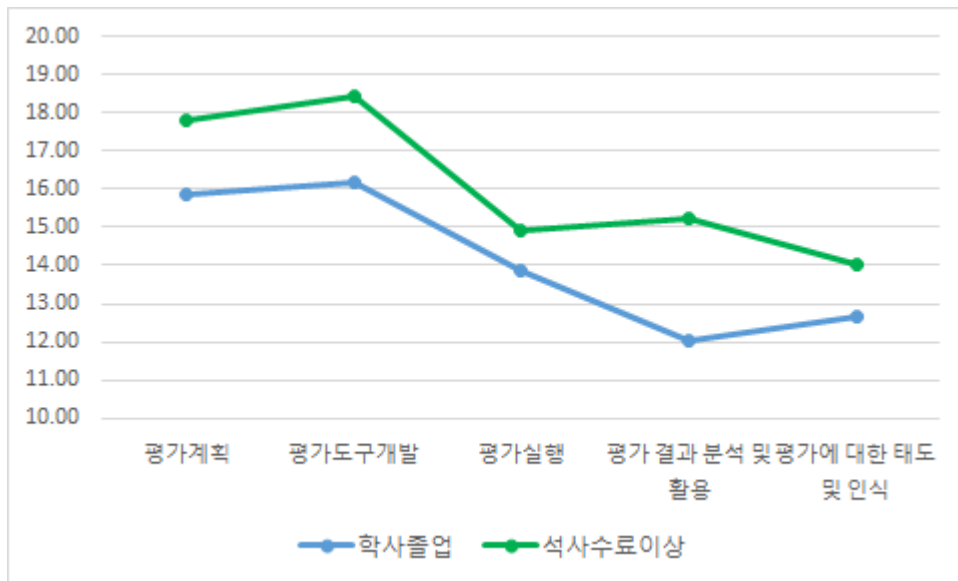
3) 학력에 따른 학생평가 전문성 분석 결과

연구에 참여한 총 36명의 영어교사들을 학력에 따라 분류하면, 학사 이상의 교사가 22명, 석사과정을 수료한 교사가 2명, 석사 졸업한 교사의 경우 12명이었다. 학력에 따른 교사의 학생평가 전문성의 기술통계 분석 결과를 살펴보면 <표 IV-8>과 같다.

<표 IV-8> 학력에 따른 학생평가 전문성의 기술통계 결과

구성요소	학사졸업(N=22)		석사수료 이상(N=14)	
	평균	표준 편차	평균	표준 편차
평가계획	15.86	2.82	17.81	3.50
평가도구개발	16.20	3.41	18.45	2.36
평가 실행	13.86	3.55	14.93	4.06
평가결과 분석 및 활용	12.03	2.95	15.21	2.95
평가에 대한 태도 및 인식	12.67	3.27	14.00	4.31
총점	70.62	11.67	80.40	13.91

기술통계 분석 결과, 학생평가 전문성의 모든 하위 영역에서 학사 졸업한 집단에 비해 석사 수료 이상의 고학력 집단의 평균 점수가 더 높은 것으로 나타났다. 또한, 학사 졸업한 교사 집단의 경우 ‘평가 실행’ 과 ‘평가도구 개발’ 영역의 표준편차가 큰 편으로 해당 영역에서 교사별 학생평가 전문성의 차이가 상대적으로 크게 나타났으며 석사수료 이상의 교사 집단에서는 ‘평가에 대한 태도 및 인식’ 영역과 ‘평가 실행’ 영역에서 비교적 학생평가 전문성의 차이가 크게 나타났다. 학사 졸업한 교사와 석사 수료 이상의 학력을 지닌 교사들 간의 학생평가 전문성의 각 하위 영역별 평균 점수를 비교하면 [그림 IV-9]와 같다.



[그림 IV-9] 학력에 따른 교사의 학생평가 전문성 평균점수 비교

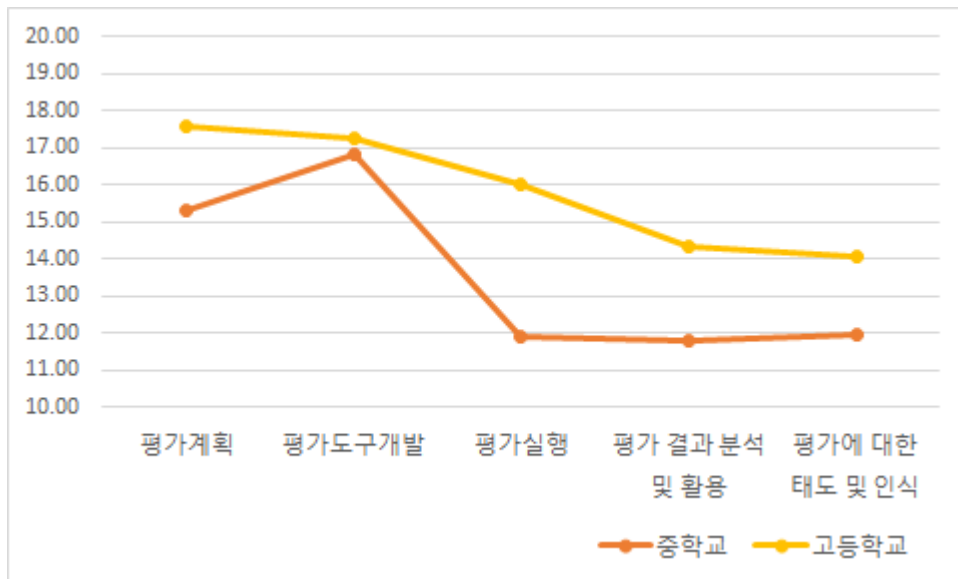
4) 학교급에 따른 학생평가 전문성 분석 결과

다음으로, 교사 전문성은 교사 개인적 특성 뿐만 아니라 학교 특성에 영향을 받는다는 선행연구(허은정, 2011)를 바탕으로 학교급에 따른 교사의 학생평가 전문성을 분석하였다. 연구에 참여한 교사 36명을 근무하는 학교급에 따라 분류해 보면, 중학교 교사가 15명, 고등학교 교사가 21명이었다. 학교급에 따른 교사의 학생평가 전문성의 기술통계 분석 결과를 살펴보면 <표 IV-9>와 같다.

〈표 IV-9〉 학교급에 따른 학생평가 전문성의 기술통계 결과

구성요소	중학교(N=15)		고등학교(N=21)	
	평균	표준편차	평균	표준편차
평가계획	15.31	3.24	17.56	2.90
평가도구개발	16.84	2.84	17.24	3.51
평가 실행	11.89	2.89	15.98	3.36
평가결과 분석 및 활용	11.80	2.91	14.32	3.23
평가에 대한 태도 및 인식	11.98	3.98	14.05	3.33
총점	67.82	10.48	79.14	13.30

기술통계 분석 결과, 모든 하위 영역에서 고등학교에서 근무하는 교사 집단의 학생평가 전문성의 평균 점수가 중학교에 근무하는 교사 집단에 비해 높게 나타났다. ‘평가도구 개발’에서는 두 집단의 평균 점수에 큰 차이가 없었으나 ‘평가 실행’ 및 ‘평가 결과 분석 및 활용’ 영역에서는 평균 점수의 차이가 두드러졌다. 중학교에서 근무하는 교사들의 경우 ‘평가에 대한 태도 및 인식’ 영역에서 교사별 학생평가 전문성의 차이가 가장 크게 나타났으며 고등학교에서 근무하는 교사들의 경우 ‘평가도구개발’ 영역의 표준편차가 상대적으로 큰 편으로 해당 영역에서 교사별 차이가 비교적 크게 나타났다. 학교급에 따른 교사의 학생평가 전문성의 각 하위 영역별 평균 점수를 비교하면 [그림 IV-10]과 같다.



[그림 IV-10] 학교급에 따른 교사의 학생평가 전문성 평균점수 비교

2.2. 평가도구의 타당도 분석

평가도구의 타당도(Validity)는 평가하고자 의도했던 구체적인 목표나 내용을 제대로 평가하고 있는가를 나타내는 정도를 의미한다(백순근, 2001). 이 연구에서는 해당 학생평가 포트폴리오 자료를 활용하여 개발한 평가척도의 각 문항이 학생평가 전문성의 구성 요인을 측정하는데 적절한지를 확인하기 위하여 평가척도의 내용 타당도에 관한 분석을 실시하였다. 다시 말해, 최종적으로 개발된 평가척도의 25개 문항이 교사의 학생평가 전문성의 5개 하위 구성요소를 측정하는데 적합한지를 확인하였다. 설문조사를 통해 영어교육 전공 교수 1인과 시간강사 1인, 박사수료 이상의 현직 영어교사 3인으로 구성된 5인의 영어 교육 전문가 집단은 평가문항의 적절성에 관해 Likert형 5점 척도(1: 전혀 적합하지

않다, 2: 적합하지 않다, 3: 보통이다, 4:적합하다, 5: 매우 적합하다)로 응답하였으며 그 결과를 제시하면 <표 IV-10>과 같다.

<표 IV-10> 문항의 적절성에 관한 내용 타당도 분석 결과

구성요인	세부 평가 기준	평가 문항의 적절성	
		평균	표준편차
평가계획	교육과정	5.00	0.00
	평가목표	5.00	0.00
	평가의 실제성	3.60	0.55
	학습자의 인지적 특성	4.80	0.45
	학습자의 정의적 특성	4.80	0.45
평가도구개발	평가도구이해	4.40	0.55
	평가도구개발	5.00	0.00
	평가도구선택	4.60	0.55
	평가도구관리	4.20	0.84
	채점기준	4.80	0.45
평가실행	평가시기	4.80	0.45
	학생안내	4.60	0.55
	채점 및 성적부여	5.00	0.00
	숙제 및 형성평가	4.60	0.55
	평가참여	4.40	0.55
평가 결과 분석 및 활용	평가 결과 분석 및 해석	4.80	0.45
	평가 결과 활용	5.00	0.00
	학부모와의 의사소통	4.20	0.45
	학생 피드백	5.00	0.00
	교수기술 및 전략	5.00	0.00
평가에 대한 태도 및 인식	평가의 공정성	4.80	0.45
	평가 활동에 대한 반성	5.00	0.00
	전문적 자율성	4.80	0.45
	동료 교사와의 협업 능력	4.00	0.71
	평가의 환류적 기능	4.00	0.71

각 세부 평가 기준이 해당 구성요소에 포함되는 것이 적절한지에 관한 전문가 평정 결과, ‘평가의 실제성’ 항목을 제외한 모든 문항의 응답 평균은 4.0이상으로 대체적으로 평가 문항이 적합한 편으로 나타났다. 반면에, ‘평가 계획’ 영역의 ‘평가의 실제성’ 항목의 평정 점수는 3.6점으로 상대적으로 낮은 편이었다. 이러한 결과는 일부 전문가들의 의견을 바탕으로 할 때 ‘평가의 실제성’ 항목의 경우 학생평가 전문성의 ‘평가 계획’ 뿐만 아니라 ‘평가 실행’ 영역과도 관련이 있기 때문으로 해석할 수 있다.

중등학교 영어교사의 학생평가 전문성 평가척도의 내용 타당도를 분석한 결과, 25개의 세부 평가 기준이 교사의 학생평가 전문성의 5가지 구성요인을 평가하기에 적절한 내용으로 구성되었음을 확인하였다. 내용 타당도 검증 과정에서 제시된 전문가들의 의견을 반영하여 최종적으로 평가도구를 수정·보완하였다.

2.3. 평가도구의 신뢰도 분석

이 연구에서는 고전검사이론과 일반화가능도이론을 각각 활용하여 교사의 학생평가 전문성 평가척도의 신뢰도를 분석하였다.

가. 고전검사이론 분석 결과

먼저, 고전검사이론을 활용하여 문항 내적 신뢰도 및 채점자간 일치도를 살펴보았다. 교사의 학생평가 전문성은 다섯 가지 하위 영역별로 5

개 문항으로 평가되었으며 Cronbach 알파 계수를 산출한 결과는 <표 IV-11>과 같다.

<표 IV-11> 평가도구의 문항 내적 신뢰도

구분	Cronbach's α	문항수
평가계획	0.67	5
평가도구개발	0.62	5
평가실행	0.56	5
평가 결과 분석 및 활용	0.63	5
평가에 대한 태도 및 인식	0.78	5
전체총점	0.88	25

분석 결과, 평가도구의 문항 내적 신뢰도는 하위 영역별로 .56~ .78로 ‘평가 실행’ 영역의 신뢰도 계수가 조금 낮은 편이었으나 전체적으로 .88로 적은 문항수를 고려했을 때 비교적 양호한 수준이었다.

다음으로, 학생평가 포트폴리오를 활용한 전문가 평가자 3인의 평정 결과에 대한 평정자간 일치도를 분석한 결과는 <표 IV-12>와 같다.

<표 IV-12> 평정 결과의 채점자간 일치도

	평가자 F						평가자 G					
	C1	C2	C3	C4	C5	총점	C1	C2	C3	C4	C5	총점
평가자 G	.86	.91	.95	.88	.93	.98						
평가자 H	.84	.89	.94	.90	.95	.97	.92	.94	.98	.94	.96	.98

이 연구에서는 특히 전문가 평가 과정에 중점을 두어 사전 채점자 모

의 훈련과 전문가 토의 과정에 많은 시간을 할애하여 채점자간 특성 차이를 최소화하고 채점자 내 일치도를 높이려고 노력하였다. 그 결과, 평가 도구의 각 하위 영역별 점수의 단순 상관계수는 .84~.98이었으며 총점의 경우 .97~.98로 높은 수준의 채점자간 일치도를 보였다.

나. 일반화가능도이론 분석 결과

다음으로, 고전검사이론에서는 확인하기 어려웠던 교사의 학생평가 전문성 평가에 영향을 주는 오차 점수의 다양한 원천을 살펴보고자 일반화가능도이론을 적용하여 평가척도의 신뢰도를 분석하였다. 허용 가능한 관찰전집과 각 분산 성분 추정치의 상대적 크기를 비교하는 G 연구를 실시하고 G 연구 결과를 바탕으로 최적의 측정 조건을 추정하는 D 연구를 실시하였다. 이 연구에서는 3명의 채점자(r)가 36명의 동일한 피험자(p)의 학생평가 전문성을 각 평가 문항(i)을 활용하여 평가하였으므로 $p \times i \times r$ 설계를 적용하여 G 연구 분석을 실시하였다. G 연구를 통해서 다른 오차원에 관한 상대적인 크기를 비교한 후 D 연구를 실시하여 G 계수 및 D 계수를 산출 한 후 신뢰도 계수가 최적화되는 측정 조건을 제시하였다.

1) G 연구 분석 결과

교사의 학생평가 전문성의 G 연구 분석 결과, 전집점수 분산과 각 오차원의 분산 성분의 크기를 상대적으로 비교하여 제시하면 <표 IV

-13>과 같다.

<표 IV-13> p x i x r 설계에 대한 G 연구 분석 결과

오차원	자유도(df)	제곱합(SS)	평균제곱(MS)	분산성분 추정값(EVC, %)
p	35	743.5630	21.2447	0.2477(12.48)
r	2	0.3563	0.1782	0.0000(0.00)
i	24	1866.4651	77.7694	0.6891(34.72)
pr	70	15.8570	0.2265	0.0005(0.03)
pi	840	2232.5482	2.6578	0.8148(41.05)
ri	48	43.1437	0.8988	0.0190(0.96)
pri,e	1680	358.6430	0.2135	0.2135(10.76)
전체	2699	5260.5762	103.1888	1.9846(100.00)

<표 IV-13>은 각 요인이 관찰 점수 분산에 미치는 영향력을 설명하고 있다. G 연구에서 분산 성분은 피험자($\hat{\sigma}^2(p)$), 채점자($\hat{\sigma}^2(r)$), 평가 문항 분산($\hat{\sigma}^2(i)$) 및 피험자와 채점자($\hat{\sigma}^2(pr)$), 피험자와 평가 문항($\hat{\sigma}^2(pi)$), 채점자와 평가 문항($\hat{\sigma}^2(ri)$) 간의 상호작용 분산과 잔차 분산($\hat{\sigma}^2(pri)$)으로 나눌 수 있다. 구체적으로 살펴보면, 총 분산($\hat{\sigma}^2(X)=1.9846$)에 대한 전집 점수 분산($\hat{\sigma}^2(p)=0.2477$)은 12.48%로 학생평가 전문성에 교사별 차이가 어느 정도 존재함을 보여주었다. 평가 문항의 차이를 나타내는 분산($\hat{\sigma}^2(i)$)과 피험자와 평가 문항($\hat{\sigma}^2(pi)$)간의 상호작용 분산은 각각 34.72%와 41.05%로 상대적으로 크게 나타났다. 이는 평가 문항이 학생평가 전문성 평가 결과에 주는 변동 요인이 크다는 것을 의미한다. 다음으로, 잔차(residual) 분산이 전체 분산의 10.76%를 차지하였다. 잔차 분산이 비교적 크게 나타나는 것은 일반화가능도 연구의 일반적인 현상으로 분산 성분 추정치에는 연구 모형으로 구분되는 효과들에 의해 설

명되지 않는 모든 잔차 분산이 포함되기 때문이다(강애남, 이규민, 2006; 박혜정 외, 2006; 양지승, 이규민, 2007).

이와는 대조적으로, 채점자 간 차이를 나타내는 분산(σ_r^2)의 경우는 0%로 평정한 채점자 간 오차 분산이 거의 존재하지 않음을 알 수 있다. 또한, 피험자와 채점자의 상호작용 분산($\hat{\sigma}^2(pr)$)과 채점자와 평가 문항의 상호작용 분산($\hat{\sigma}^2(ri)$)은 각각 0.03%와 0.96%로 매우 작게 나타났다. 이는 실제 분석 결과 높은 채점자 간 일치도에서 알 수 있듯이 채점자 간 점수 차이가 크지 않았기 때문으로 유추할 수 있다.

2) D 연구 분석 결과

D 연구에서는 G 연구에서 산출된 분산성분 추정 값을 바탕으로 각 오차 국면의 조건 수를 변화시킴으로써 G 계수와 준거 참조 평가를 위한 D 계수의 변화를 분석하였다. 또한, 교사의 학생평가 전문성을 평가할 때 적정 수준의 신뢰도를 확보하기 위한 효율적인 측정조건을 모색하였다. D 연구의 분석 결과를 정리하면 <표 IV-14>와 같다.

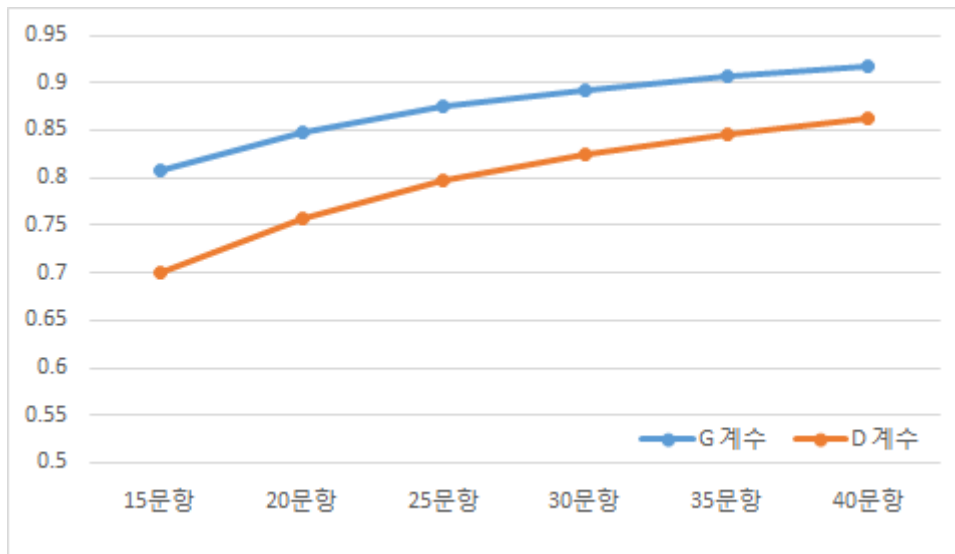
<표 IV-14> p X I X R 설계에 대한 D 연구 분석 결과

조정 국면 (facet)	측정 조건		전집 점수	관찰점수 기대치	상대 오차	절대 오차	G 계수	D 계수
	채점자	문항						
채점자	2	25	0.2477	0.2848	0.0371	0.0651	0.8697	0.7919
	3	25	0.2477	0.2833	0.0356	0.0634	0.8743	0.7961
	4	25	0.2477	0.2825	0.0349	0.0626	0.8766	0.7982
	5	25	0.2477	0.2821	0.0344	0.0621	0.8780	0.7995
	6	25	0.2477	0.2818	0.0341	0.0618	0.8790	0.8003
문항	3	15	0.2477	0.3069	0.0592	0.1056	0.8070	0.7011
	3	20	0.2477	0.2921	0.0445	0.0792	0.8478	0.7576
	3	30	0.2477	0.2774	0.0297	0.0529	0.8929	0.8240
	3	35	0.2477	0.2731	0.0255	0.0454	0.9067	0.8452
	3	40	0.2477	0.2700	0.0223	0.0397	0.9173	0.8618

D 연구 설계에서 가장 중요한 것은 일반화 전집(universe of generalization)을 규정하는 것이다(김성숙, 김양분, 2001). 이 연구에서 일반화 전집은 채점자 국면 및 평가 문항 국면이며 모두 무선효과를 가정하였고, 관련 선행연구 검토를 바탕으로 교사의 학생평가 전문성을 구성하는 하위 영역 국면을 5개로 고정하였다. 또한, 상대오차 분산 및 절대오차 분산을 산출하여 각각 G 계수와 D 계수를 산출하였다. 전반적으로 상대오차 분산의 크기가 절대오차 분산의 크기보다 작게 추정되어 상대오차 분산을 활용한 G 계수의 크기가 D 계수보다 크게 산출되었다. 분석 결과, 이 연구의 측정 조건(3명의 채점자, 25개의 평가 문항)에서 도출된 G 계수는 .8743이며 D 계수는 .7961로 G 계수가 .8 이상, D 계수가 .7이상으로 적절한 수준으로 나타났다.

다음으로 각 오차 국면을 조정하여 G 계수와 D 계수의 변화를 살펴 보았다. 채점자 국면의 경우, G 연구 결과에서도 알 수 있듯이 채점자에

의한 오차 분산이 거의 존재하지 않아 채점자를 2명 ~ 6명으로 증가시킴에 따라 신뢰도 계수가 크게 변화하지 않았다. 이와는 대조적으로, 평가 문항 수 국면의 경우, 문항의 수를 점차 늘림에 따라서 G 계수와 D 계수가 증가하고 있음을 알 수 있다. D 연구에서 문항 수의 변화에 따른 G 계수와 D 계수의 변화를 구체적으로 살펴보면, [그림 IV-11]과 같다.



[IV-11] 평가 문항수에 따른 G 계수 및 D 계수의 변화 양상

D 연구 분석 결과, 채점자 수가 동일한 조건일 경우, 평가 문항의 수가 증가함에 따라서 G 계수와 D 계수 또한 증가하는 것으로 나타났다. 구체적으로 살펴보면, 채점자가 3명일 때 평가 문항의 수가 증가할수록 신뢰도 계수가 점점 높아지는 경향이 있으며 평가 문항이 15개인 경우 이미 적절한 수준의 신뢰도를 확보하는 것으로 나타났다. 이는 간편형 평가척도 개발 시 적절한 신뢰도를 확보할 수 있는 평가 문항의 수를 의

미한다고 할 수 있다.

반면에, 평가 문항의 수가 동일한 조건일 경우, 채점자 수에 따른 신뢰도 계수의 증가폭이 매우 작은 것을 알 수 있다. 즉, 채점자 수를 늘리는 것보다 평가 문항의 수를 증가시키는 것이 연구에서 설정한 신뢰도를 확보하는데 더욱 효과적임을 알 수 있다. 따라서, 적절한 수준의 신뢰도 확보 뿐만 아니라 효율적인 측정을 위해 채점자 수는 2명, 평가 척도의 문항 수는 25개가 적절하다고 판단되었다.

2.4. 평가도구의 적합도 분석

연구에 참여한 중등학교 영어교사와 학생평가 포트폴리오의 전문가 평가자를 대상으로 실시한 개방형 설문조사를 바탕으로 평가도구의 적합도에 대해 분석한 결과를 정리하면 다음과 같다.

첫째, 연구 참여 교사들의 설문 응답을 분석한 결과, 다수의 교사들은 학생평가 포트폴리오 평가를 통해 자신의 학생평가 활동 전반에 대해 성찰하고 반성할 수 있는 기회를 가질 수 있었으며, 형식적인 학생평가 활동 뿐만 아니라 실제 교수·학습 과정과 통합된 비형식적 학생평가 활동의 중요성을 깨닫게 되었다고 응답하였다. 이는 2015 개정교육과정에서 강조하고 있는 ‘과정중심평가’에 학생평가 포트폴리오 평가가 유용하게 활용될 수 있는 가능성을 의미한다. 이와 관련된 설문 내용의 일부를 제시하면 다음과 같다.

“한 학기의 모든 평가 항목들과 평가 과정을 돌아보며, 계획 수립에서부터 평가 결과에 이르기까지 스스로의 학생평가 활동을 평가하고 반성

할 수 있었습니다. 또한, 학생평가 활동과 관련된 자료들을 수집하는 일련의 과정에서 많은 것을 배울 수 있었습니다. 학기 초에 세운 평가 계획들을 충실히 이행하고 반영했는지, 공정한 평가를 위해 노력을 했는지, 평가 결과물이 타당한지에 대해 스스로 질문해 볼 수 있었고, 부족한 점이나 잘못된 점에 대해 생각해 볼 수 있었습니다. 특히, 이 과정에서 공식적 평가 결과뿐 아니라, 수업 시간에 개인적으로 실시했던 비공식적 학생평가의 중요성을 깨달았고, 학생들의 개인적 노력과 특성을 관찰하고 기록하여 최종적으로 이를 생활기록부의 세부 특기 사항에 잘 반영해 주는 것이 학생들 개인적으로도 의미가 있음을 알게 되었습니다. 아울러 영어 수업 시간에 학생들의 수준 차이를 반영하고 모둠별 활동 기회를 제공하는 것이 학생들의 수업 참여도를 높여주는 것임을 확인할 수 있었습니다. 이는 교사 입장에서 학생들의 개별적 특성을 파악하고 학생들을 위한 개별적 평가에도 유용했습니다. 즉, 수업과 평가가 동떨어진 것이 아니라, 다양한 수행평가 및 지필평가를 대비할 수 있는 평소 수업 활동의 제공이 중요하다는 것을 실감하였습니다. ‘학생평가 포트폴리오 평가’는 저에게 학생평가 활동과 관련된 자기반성의 경험을 제공하였으며 이를 통해 앞으로 보다 주도적인 입장에서 다양한 공식적, 비공식적 평가를 실시해야 함을 깨닫게 되었습니다(교사 4).”

“학생평가 포트폴리오 평가를 통해서 스스로의 학생평가 활동에 있어서 큰 그림을 그릴 수 있게 되었습니다. 교육 과정 및 수업 계획에서부터 각종 평가의 역할 및 성적 처리, 그리고 평가에 대한 교사 전문성 향상의 필요성에 이르기까지 결국 수업도 큰 의미에서 바라보면 평가의 일부가 될 수 있다고 느꼈습니다. 즉, 평가는 단지 학생이 수행한 결과물에 대해 성적을 부여하는 것이 아니라 교사가 수업의 내용 및 평가하고자 하는 요소에 대한 분명한 인식을 바탕으로 학생들의 수행을 관찰하고 행동에 대해 피드백을 주어 목표한 결과를 달성하도록 이끌어 가는 과정 전체를 의미하는 것입니다. 이를 위해 교사는 학생평가 전문성을 향상시키기 위

해 지속적으로 노력해야 함을 깨달았습니다(교사 8).”

둘째, 학생평가 포트폴리오 평가는 기존의 현직 교사 연수와 비교했을 때 학생평가 전문성의 실천적 역량을 신장시키는데 도움을 준 것으로 나타났다. 일부 교사들의 경우 평가 전문성 관련 연수에 참여한 경험이야에 없거나 참여한 경우에도 원론적인 내용을 전달하는 연수에 그쳐 학교 현장에서 평가 전문성 향상을 위해 실질적으로 참고하거나 활용할 만한 내용은 부족하였다고 응답하였다. 또한, 교사마다 교수·학습 환경이 상이하여 평가 연수를 통해 배운 내용을 실제 학교 현장에서 적용하는데 어려움을 겪는 것으로 나타났다. 이러한 측면에서, 학생평가 포트폴리오 평가를 교사의 학생평가 전문성 향상을 위한 보다 실제적이고 효과적인 방안으로 제시할 수 있을 것이라는 의견이 있었다.

“지금까지 참여한 평가 관련 교사 연수중에서 실제 학교 현장에서 이루어지는 사례를 공유하거나 직접 평가도구를 개발해보는 과정에 참여하는 방식의 연수는 실제 평가 전문성 신장에 많은 도움이 되었습니다. 하지만, 평가 관련 연수 후 교사 연수를 통해 배운 것을 학교 현장에서 그대로 적용하는 것은 쉽지 않았습니다. 연수 때 배운 것을 바탕으로 제가 가르치는 학생들과 학교 상황에 맞게 최적화된 수업을 계획하고 평가를 해야 하기 때문에 다른 교사가 사용한 평가도구를 그대로 적용하기 어려웠습니다. 이런 상황을 비추어 볼 때, 이번에 참여한 ‘학생평가 포트폴리오 평가’는 학생평가 전문성을 실제로 향상시키는데 도움이 되었다고 생각합니다. 다양한 평가 이론을 이해하고 동료 교사들의 평가 방식을 공유하는 것에서 나아가 스스로의 학생평가 활동 전반에 대해 반성해 볼 수 있었고, 평가 활동을 더욱 큰 교육 과정의 그림 속에서 파악하며 그 중요성을 더 깊이 깨닫게 되었기 때문입니다. 따라서, 더욱 많은 교사들

이 학생평가 포트폴리오를 활용할 필요가 있으며 효율적인 활용을 위한 교육청 차원의 적절한 지원이 뒷받침 되어야 할 것입니다. 교사는 학생평가 포트폴리오 제작 과정을 통해 스스로의 학생평가 활동에 대한 유의미한 깨달음을 얻게 될 뿐만 아니라, 나아가 이러한 노력은 자연스럽게 교사의 학생평가 전문성 향상으로 이어질 것으로 기대됩니다(교사 11).”

앞서 살펴본 바와 같이 학생평가 포트폴리오를 활용한 평가는 교사의 전반적인 학생평가 활동을 성찰해보고 실제적인 평가 전문성을 신장시키는데 유용한 방안으로 활용될 수 있다.

반면, 연구 참여 교사들과 전문가 평가자들의 응답 분석 결과, 학생평가 포트폴리오 평가의 한계점으로 지적된 내용을 요약하면 다음과 같다. 먼저 많은 교사들은 업무 가중으로 인한 심리적인 부담감과 학생평가 포트폴리오의 데이터 수집상의 어려움을 느끼는 것으로 나타났다.

“학생평가 활동을 수행할 때 마다 평가 문항과 채점기준, 학생 답안 샘플 등을 미리 모아놓지 않았습니다. 학생평가 포트폴리오의 구성요소를 미리 점검하지 않고 한꺼번에 수집하느라 힘들었습니다(교사 21).”

“학생들의 평소 수업 태도 및 참여 정도를 구체적인 기록으로 남겨놓지 않아서 과제 및 학습 태도 체크리스트 작성에 어려움을 느꼈습니다(교사 27).”

특히, 많은 교사들은 공통적으로 학생평가 포트폴리오의 데이터 수집 과정에서 평가 후 교사의 피드백 활동을 제대로 실시하지 않아 어려움을 겪은 것으로 나타났다.

“수행평가 항목이 여러 개이고, 7개 학급을 대상으로 실시했던 다양한 모둠별 활동, 리딩 로그 작성 및 서술형 수행평가 결과에 대해 수업 시간마다 모든 학생에게 개인적 피드백을 제공하기 힘들었습니다. 또한, 주로 개인적으로 문의한 학생들만을 대상으로 피드백을 제공하였기 때문에 자료가 남아있지 않는 경우도 많았습니다(교사 3).”

“평가 후 교사의 피드백 항목에 대한 자료를 수집하면서 평소 학생들에게 직접적으로 피드백을 제공한 경험이 부족한 것을 알게 되었습니다. 성적에 정식으로 반영되는 수행평가 작문 활동은 개별 채점과 피드백이 비교적 자세히 제공되었으나, 평소 수업시간에 이루어지는 퀴즈나 형성평가 및 과제에 대한 피드백의 경우 학생들이 작성한 결과물을 교사가 모두 검토하기에는 담당하고 있는 학생들의 수가 너무 많아 일괄적인 검토나 채점 또는 단순한 과제 확인 정도로 피드백을 제공하였습니다(교사 6).”

“예전 신규 교사일 때는 기말고사가 끝나고 나서 개별 학생마다 평가 결과에 관한 분석과 의견을 자세하게 기록하여 각 학부모에게 가정 통신문 형태로 보낸 적이 있습니다. 현재 근무하는 학교의 경우에는 업무가 너무 많고 따로 기록으로 남기지 않아 ‘평가 후 교사의 피드백’ 자료를 수집하기 어려웠습니다(교사14).”

다음으로, 3인의 전문가 평가자들의 의견을 바탕으로 평가도구의 개선할 점을 살펴보면 다음과 같다. 첫째, 전문가 평가자들은 학생평가 포트폴리오만을 평가 자료로 활용하여 교사의 학생평가 전문성을 다각도로 평가해야 하는 점에서 어려움을 느낀 것으로 응답하였다. 또한, 학생평가 포트폴리오에는 명시적으로 제시되지 않으나 학교 교육 과정이나 눈에 보이지 않는 교사 특성이 교사의 학생평가 전문성에 영향을 줄 수 있

음을 지적하였다. 전문가 평가자들의 구체적인 설문 내용의 일부를 제시하면 다음과 같다.

“학생평가 포트폴리오를 활용해서 교사의 학생평가 전문성을 평가할 때, 교사의 실제 평가 전문성 뿐만 아니라 자료 수집 및 정리에 있어서의 성실성, 꼼꼼함 등이 추가로 평정 결과에 영향을 미치는 것 같다는 생각이 들었습니다. 평정자는 학생평가 포트폴리오에 제시된 내용만 가지고 유추하여 평가해야 하기 때문에, 교사의 학생평가 전문성이 실제로 높고 교육 현장에서는 제대로 수행한 부분임에도 포트폴리오에 자료가 충분하지 않다면 낮은 점수를 부여할 수 밖에 없습니다. 포트폴리오 자료를 통해 ‘간접적’으로 교사의 학생평가 전문성을 평가해야 하기 때문에 나타나는 어쩔 수 없는 한계점이라고 생각되지만, 일부 교사들의 경우 실제 평가 전문성이 낮은 교사들이라기보다 자료 수집이 꼼꼼하지 않은 교사일 수 있다는 점에서 평가도구의 타당도를 낮추는 요소가 될 수도 있을 것 같습니다. 가능하다면 이러한 부분을 보완할 수 있는 방안에 대한 고민이 필요할 것으로 생각합니다(전문가 F).”

둘째, 전문가 평가자들은 공통적으로 평가도구의 필요성과 적합성에 대해서는 공감하였으나, 일부 평가 기준의 경우 실제 교육 현장을 반영하지 못하는 점을 지적하였다. 구체적인 평가 기준을 예로 들어 살펴보면 다음과 같다.

“전체적으로 평가 전문성의 다양한 부분을 종합적으로 잘 포괄하고 있는 척도라고 생각합니다. 단지, 실제로 어려울 수 있는 부분이라 생각되었던 부분은 평가척도 8번 문항에서 “교사는 평가 목적에 따라 기존에 개발된 다양한 유형의 표준화 검사를 선택하고 활용할 수 있는가?”를 묻고 있는데, 영어 교과에 한정해서 활용할 수 있는 표준화 검사가 EBS

듣기평가 등으로 한정되어 있기 때문에 교사가 평가도구를 ‘선택’ 하는 측면에서 한계가 있을 것 같습니다. 또한, 실제로 평정을 해 보았을 때 평가척도 15번의 ‘평가참여’, 18번의 ‘학부모와의 의사소통’ 문항의 경우, 높은 점수를 부여할 수 있는 대상이 많지 않았고, 평가척도 13번의 ‘전문성 자율성’의 경우에도 채점기준에 따라 평정하였을 때 4점 이상의 높은 점수를 부여할 수 있는 대상이 거의 없었습니다. 따라서, 이 기준이 교육 현장에서는 실제적으로 충족하기 어려운 기준이 아닌지 면밀하게 검토해 볼 필요가 있을 것 같습니다(전문가 F).”

“평가척도 23번의 ‘전문적 자율성’의 채점기준에서 단순히 교사 교육 참여 횟수만이 아니라 교육 참여 시간도 함께 평가하면 좀 더 교육 현실을 반영하는 평가가 될 것이라고 생각합니다(전문가 G).”

“이론상으로는 모두 필요한 평가 항목이나 평가척도 8번 문항의 ‘평가 도구 선택’에서 표준화 검사를 활용하고 선택하는지 여부는 교육 현실상 학교나 지역 교육청의 방침을 따를 수 밖에 없습니다. 평가척도 18번 문항의 ‘학부모와의 의사소통’의 경우에도 중등 교육 과정의 특성상 교과 교사가 학부모와의 의사소통을 하는 경우가 거의 없는 현실을 반영하였을 때 실제 학교 현장에서는 적용하기 어려운 평가 항목이 아닌가 생각합니다(전문가 H).”

전문가 평가자들은 공통적으로 ‘학부모와의 의사소통능력’이나 ‘평가도구를 선택 할 수 있는 능력’이 이론적으로는 교사의 학생평가 전문성 평가 기준에 꼭 포함되어야 하는 항목이나 실제 학교 현장에서 이를 제대로 활용하는 교사가 많지 않거나 학생평가 포트폴리오 자료에 제대로 드러나지 않은 경우가 존재함을 지적하였다. 따라서, 교사들에게 해

당 평가 기준을 적용할 수 있는 보다 현실적이고 구체적인 상황을 제시하거나 교사 설문조사, 인터뷰나 수업 관찰 방법 등의 질적 평가 방법을 통해 이를 보완할 필요가 있다.

셋째, 전문가 평가자들은 학생평가 포트폴리오를 활용한 교사의 학생평가 전문성 평가에 있어서 보완되어야 할 점으로 연구 대상의 표집과 평가도구의 실용성 측면을 지적하였다.

“동일 학교에서 섭외한 몇 개의 학생평가 포트폴리오의 경우 ‘평가에 대한 태도 및 인식’ 항목을 제외한 다른 영역에서 비교적 비슷한 결과를 보이는 경향이 있었습니다. 즉, 교사 개인의 포트폴리오라고는 하지만 근무하는 학교의 교육 과정이나 동료 교사에 따라 평가도구 개발, 평가 실행 및 평가결과 분석 영역에서 교사 개인의 학생평가 역량에 보이지 않는 영향을 미칠 가능성이 있습니다. 향후 연구에서는 연구 대상을 표집할 때 보다 대표성을 띄게 할 수 있도록 다양한 학교에서 근무하는 중·고등학교 교사를 대상으로 실시할 필요가 있을 것 같습니다(전문가 G).”

“학생평가 포트폴리오를 활용한 평가의 현장 적합성을 높이기 위해서 평가 문항을 간소화한 보다 간편형 평가도구의 제작도 고려할 필요가 있다고 생각합니다(전문가 H).”

학생평가 포트폴리오 평가는 교사의 전반적인 학생평가 활동에 대해 반성 및 성찰의 기회를 제공하고 교사의 실천적 평가 전문성 신장에 도움이 줄 수 있다는 점에서 평가도구의 본질적인 목적에 적합하다는 의견이 많았다. 그럼에도 불구하고 학생평가 포트폴리오만을 평가 자료로 활용하여 교사의 학생평가 전문성을 다각도로 평가해야 하는 점이나 일부

평가 기준의 경우 실제 교육 현장을 반영하지 못하는 점 등은 한계점으로 지적되었다.

V. 요약 및 논의

1. 요약

이 연구의 목적은 중등학교 영어교사의 실천적 학생평가 전문성을 평가하기 위한 도구를 개발하고 교사의 학생평가 포트폴리오를 활용한 전문가 평가를 실시하여 개발된 평가도구의 타당도, 신뢰도 및 적합도를 확인하는 것이다. 이를 위해 설정한 연구 문제는 다음과 같다.

1. 중등학교 영어교사의 학생평가 전문성의 평가 기준은 무엇인가?
2. 중등학교 영어교사의 학생평가 전문성 평가도구의 내용 타당도는 적합한가?
3. 중등학교 영어교사의 학생평가 전문성 평가도구의 신뢰도는 적합한가?
4. 중등학교 영어교사의 학생평가 전문성 평가도구는 평가의 목적과 취지에 적합한가?

이러한 연구 문제를 검증하기 위해 교사의 학생평가 전문성 및 교사 포트폴리오와 관련된 다양한 선행 연구 분석과 전문가 검토 과정을 거쳐 중등학교 영어교사의 학생평가 전문성 평가척도, 학생평가 포트폴리오의 구성요소 및 전문가 평정을 위한 채점기준을 각각 개발하였다.

평가도구를 개발한 다음, 서울, 경기, 인천 지역의 중·고등학교에서 근무하는 영어교사 42명을 대상으로 한 학기(2016년 2학기) 동안의 학

생평가 활동과 관련된 학생평가 포트폴리오를 수집하였다. 그 중 36명의 영어교사들의 실제 데이터를 바탕으로 개발된 평가도구를 활용하여 교육 측정 및 평가 분야의 전문가 1인과 영어 교육 분야의 평가 전문가 2인으로 구성된 3인의 전문가 평가단이 학생평가 포트폴리오에 대한 평가를 실시하였다. 전문가 평가 결과를 바탕으로 평가척도의 세부 평가 기준 및 교사 배경 변수에 따른 교사의 학생평가 전문성의 구체적인 특성을 분석하였으며, 평가도구의 타당도와 신뢰도에 관해 검증하였다. 최종적으로 연구 참여 교사 및 전문가 평가자를 대상으로 평가도구의 적합도 및 학생평가 포트폴리오 평가 참여 소감에 관한 개방형 설문조사를 실시하였다.

주요 연구 결과를 요약하면 다음과 같다. 첫째, 중등학교 영어교사의 학생평가 전문성 평가척도는 ‘평가 계획’, ‘평가도구 개발’, ‘평가 실행’, ‘평가 결과 분석 및 활용’, ‘평가에 대한 태도 및 인식’의 5개의 하위 영역으로 구성되었고 각 영역별로 5개의 세부적인 평가 기준을 포함하였다. 하위 영역별로 학생평가 전문성의 평균 점수를 비교해보면, ‘평가도구 개발’ 영역이 가장 높게 나타났으며, ‘평가 결과 분석 및 활용’이나 ‘평가에 대한 태도 및 인식’ 영역의 평균 점수는 상대적으로 낮은 편이었다. 세부 평가 기준별로 살펴보면, ‘평가의 실제성’ 항목의 점수가 가장 높은 반면 ‘평가 참여’와 ‘학부모와의 의사소통’ 항목의 평균 점수가 가장 낮은 것으로 나타났다.

둘째, 영어교육 전공 교수 1인과 시간강사 1인, 박사 수료 이상의 현직 영어교사 3인을 포함하여 총 5인으로 구성된 영어교육 전문가 집단을 대상으로 개발한 평가도구의 내용타당도에 관한 검증을 실시하였다. 각 문항이 해당 학생평가 포트폴리오를 평가 자료로 활용하여 학생평가

전문성의 하위 구성 요인을 측정하는데 적절한지에 관한 전문가 응답 분석 결과, 25개의 세부 평가 기준은 교사의 학생평가 전문성의 다섯 가지 구성요인을 평가하기에 적절한 내용으로 구성되었음을 확인할 수 있었다.

셋째, 평가도구의 신뢰도를 검증한 결과, 비교적 양호한 수준의 신뢰도를 확보한 것으로 나타났다. 구체적으로 살펴보면, 고전검사이론을 활용하여 평가 문항의 내적 신뢰도를 분석한 결과, 하위 영역별로 Cronbach의 알파계수가 .56~.78이었으며, 총점의 경우 .88을 나타냈다. 또한, 전문가 3인의 평정 결과 간의 단순 상관계수는 각 영역별로 .84~.98로 나타나 높은 수준의 채점자 간 일치도를 보여주었다.

다음으로, 일반화가능도이론을 활용하여 신뢰도에 영향을 미치는 다양한 오차원과 그 상대적인 영향력을 분석하였으며, 안정적인 신뢰도 계수를 얻기 위한 효율적인 측정 조건을 탐색하였다. G 연구 결과 평가 문항의 차이를 나타내는 분산($\hat{\sigma}^2(i)$)이 오차 분산 중 상대적으로 큰 것(34.72%)으로 나타나 평가 문항이 학생평가 전문성의 평가 결과에 주는 변동 요인이 크다는 것을 알 수 있다. 피험자 간 차이를 의미하는 분산($\hat{\sigma}^2(p)$)의 경우 12.48%를 차지해 학생평가 전문성에 교사별 차이가 어느 정도 존재함을 보여주었다. 반면에, 채점자 간 오차분산은 거의 존재하지 않았다.

D 연구 분석 결과를 살펴보면, 이 연구의 측정 조건 (3명의 채점자, 25개의 평가 문항)에서 도출된 G 계수는 .8743이며 D 계수는 .7961로 각각 .8과 .7 이상으로 적절한 수준의 신뢰도를 확보한 것으로 나타났다. 이 연구에서 설계한 오차 국면을 조정하여 G 계수 및 D 계수의 변화를 살펴본 결과, 채점자 수에 따라 신뢰도 계수의 변화가 거의 존재하지 않은 것과는 대조적으로 평가 문항의 수를 증가시킬수록 더욱 안정적

인 측정 결과를 제시하는 것으로 나타났다. 따라서, 이 연구에서는 측정 점수의 신뢰도를 확보하기 위해 채점자 국면 보다는 평가 문항 측면에서의 고려가 필요함을 알 수 있다.

넷째, 연구 참여 교사 및 전문가 평가자들을 대상으로 실시한 개방형 설문조사의 결과를 바탕으로 평가도구의 적합도를 살펴보면 다음과 같다. 학생평가 포트폴리오 평가는 교사들에게 학생평가 활동 전반에 대한 성찰과 반성의 기회를 제공하며, 기존 평가 연수에 참여한 경험과 비교했을 때 교사의 학생평가 전문성 향상을 위한 보다 실제적이고 효과적인 방안이라는 의견이 있었다. 한편, 일부 평가 항목의 경우 데이터 수집상의 어려움이 있으며, 해당 포트폴리오 자료만을 활용하여 교사의 학생평가 전문성을 다각적인 측면에서 평가해야 하는 점은 개발된 평가도구의 한계점으로 지적되었다. 이러한 연구 결과는 학생평가 포트폴리오의 데이터 수집이나 평정 과정에서 나타나는 일부 어려움을 보완한다면 교사의 학생평가 전문성 평가도구가 교사의 실천적인 학생평가 전문성을 신장시키기 위한 교사 교육 및 교사 연수에서 보다 효과적으로 활용될 수 있음을 시사한다.

2. 논의 및 시사점

교사의 학생평가 전문성은 교실 수업의 개선과 학습자의 학습 활동에 도움을 줄 뿐만 아니라 나아가 공교육 강화를 위한 필수적인 요건이라 할 수 있다. 이 연구는 중등 영어교사의 실천적 학생평가 전문성을 평가할 수 있는 도구를 개발하고, 수집된 교사의 학생평가 포트폴리오를 활용하여 개발한 평가도구의 타당도, 신뢰도 및 적합도를 검증하였다. 연구의 과정 및 결과에서 발견된 몇 가지 논의사항을 정리하면 다음과 같다.

첫째, 하위 영역별로 교사의 학생평가 전문성을 분석한 결과, ‘평가 도구 개발’ 영역의 점수가 가장 높았으며 ‘평가 결과 분석 및 활용’ 이나 ‘평가에 대한 태도 및 인식’ 영역의 점수는 상대적으로 낮게 나타났다. 이는 오랜 기간 동안 교사의 학생평가 전문성이 대부분 수행평가나 지필평가 등의 문항 출제 능력 위주로 강조되어 온 반면, 학교 현장에서 평가 결과에 대한 충분한 분석 및 활용 능력의 중요성이 상대적으로 간과되어 왔기 때문이다. 또한, 공정한 평가, 교사의 평가 활동에 관한 반성, 평가 전문성 개발 노력, 평가 과정에서의 동료 교사와의 협업 능력 및 평가의 환류적 기능(washback) 등 평가 활동 전반에서 교사의 태도 및 인식을 상대적으로 소홀히 한 결과가 반영되었다고 할 수 있다. 이러한 연구 결과는 평가 실행 이후 ‘평가 결과에 대한 분석 및 활용’ 과 ‘평가에 대한 태도 및 인식’ 영역에 관한 교사의 학생평가 전문성에 더 많은 관심을 기울이고 이를 신장시킬 수 있는 방안에 대한 고려가 필요함을 시사한다.

둘째, 이 연구의 평가도구는 중등학교 영어교사의 실제적인 학생평가 전문성을 평가하기 위해 개발된 것이다. 이를 위해 영어과 교육과정, 영

어 교과에서의 수업평가 및 교사평가와 관련된 선행연구를 바탕으로 하여 평가척도의 예비 문항, 학생평가 포트폴리오 구성요소 및 채점기준을 개발하였다. 또한, 영어 교육 분야의 전문가 5인으로부터 평가도구의 내용 타당도에 관한 검증을 실시하였으며, 학생평가 포트폴리오 및 채점기준의 세부 내용에 영어교과에서의 학생평가 특성이 주로 반영되었다는 점에서 영역 특수적인 성격을 지닌다고 할 수 있다. 하지만, 전문가 검토 과정을 거쳐 평가척도를 수정·보완하여 최종화 하는 과정에서 과목 특수적인 평가 항목이 일부 제외되었으며 상당수의 평가 문항의 경우 영역 일반적으로 적용할 수 있는 내용을 또한 포함하고 있다. 따라서, 중등학교 영어교사의 학생평가 전문성 평가도구는 평가 대상, 학생평가 포트폴리오의 세부 내용 및 채점기준을 각 교과목의 특성에 따라 수정하여 보다 광범위하게 활용될 수 있을 것이다.

셋째, 학생평가 포트폴리오를 활용한 평가는 교사의 실천적 학생평가 전문성 신장에 도움을 주는 것으로 나타났다. 연구 참여 교사를 대상으로 실시한 개방형 설문 조사에 의하면, 대부분의 교사들이 학생평가 활동에 있어서 피드백 제공의 필요성과 중요성에 대해서는 공감하고 있으나 이를 학교 현장에서 적용하지 못하는 것을 알 수 있었다. 중등학교 영어교사의 경우, 교사 당 담당하는 학생 수가 평균 5~7개 학급으로 많은 편으로 현실적으로 교사가 모든 학생들에게 개인적인 피드백을 제공하기 어렵다는 한계점이 발견되었다. 또한, 일부 피드백을 제공한 경우에도 지필평가나 수행평가 등 공식적인 평가 활동에 국한되어 있으며, 평소 수업시간에 이루어지는 형성평가나 퀴즈, 질문, 과제에 대한 피드백의 경우는 더욱 부족한 것으로 나타났다. 따라서, 학생들에게 개별적인 피드백을 제공하는 것이 현실적으로 어렵다는 점을 감안할 때 학생들에게 동료 평가의 기회를 부여하거나 모둠별로 피드백을 제공하는 방안

을 활용할 수 있을 것이다.

또한, 평가 관련 연수에 관한 교사들의 경험을 묻는 설문 내용을 바탕으로 할 때, 많은 교사들이 평가 관련 연수에 참여한 경험이 아예 없거나 참여한 경우에도 연수 내용에 관해 불만족 한 것으로 나타났다. 대부분의 평가 연수가 평가 이론이나 지나치게 원론적인 내용을 다루는 경우가 많았으며, ‘평가도구 개발’ 영역에만 초점을 두어 평가 전문성의 다른 하위 영역을 신장시키는 데에는 도움을 주지 못하였다. 또한, 교사마다 가르치는 학생이나 소속 학교의 환경이 다르므로 교사가 평가 연수에서 배운 내용을 교육 현장에서 실제로 적용하기 어려운 점도 지적되었다. 반면에, 대규모의 교사 연수 기관에서 이루어지는 연수보다는 소규모의 교사 동아리나 전문적 학습 공동체를 통해 실제 교육 현장에서의 평가 사례를 공유하거나 동료교사들과 함께 평가도구를 직접 개발해 보는 평가 연수의 경우 비교적 효과적인 것으로 응답하였다.

이러한 측면에서 학생평가 포트폴리오를 활용한 평가는 동료 교사와의 협력을 통해 교사 자신의 평가 노하우를 공유할 수 있을 뿐만 아니라 평가 계획에서부터 평가 결과를 분석하고 활용하는 영역까지 스스로의 학생평가 활동 전반에 대한 성찰과 반성을 가능하게 하므로 교사의 실제적 평가 전문성 신장을 위해 효과적인 대안이라고 할 수 있다.

한편, 많은 교사 및 교사 교육자들이 학생평가 포트폴리오를 활용한 평가의 필요성을 인식하고 있으나 학교 현장에서 실제로 활용하는 경우는 부족한 실정이다. 이는 국가 차원의 교사 평가 기준표를 기준으로 거의 모든 교사들을 대상으로 포트폴리오 평가를 실시하고 있는 미국의 사례와는 대조적이라 할 수 있다(손영민, 2007). 교사가 스스로 포트폴리오를 자율적으로 구성하고 활용하는데 있어서 현실적인 어려움이 있음을 감안 할 때, 교육청 차원에서 학생평가 포트폴리오의 구성요소 및 자료

선별 방법에 관한 구체적인 가이드라인을 제시하거나 학생평가 포트폴리오의 개발을 지원할 필요성이 있다. 학생평가 포트폴리오 구성에 있어서 많은 시간이 소모되거나 교사가 심리적 부담감을 느끼는 등의 문제점을 극복하고 포트폴리오 평가의 지속적 활용을 위해서는 전자 포트폴리오 방식을 포함한 보다 효율적인 평가 포트폴리오의 개발이 필요할 것이다.

이 연구가 지니는 의의를 살펴보면 다음과 같다.

첫째, ‘학생평가 포트폴리오’를 평가 자료로 활용함으로써, 일회성 또는 단기간의 평가가 아닌 장기간의 실제적인(authentic) 교사의 학생평가 전문성을 평가할 수 있는 도구를 개발하였다는 점이다. 학생평가 포트폴리오를 활용한 평가는 공식적 학생평가 결과 뿐만 아니라 퀴즈, 질문, 형성평가, 과제 등 평소 수업시간에 수행되는 피드백 제공 활동이나 비공식적 평가 활동에 관한 정보를 제공하며, 수업 활동과 통합된 형태의 학생평가 활동을 수행하는데 장점이 있는 것으로 나타났다. 또한, 교사 입장에서는 ‘학생평가 포트폴리오’의 수집 과정을 통해서 평가 계획에서부터 평가도구 개발, 평가 실행, 평가 결과 분석 및 활용에 이르기까지 전반적인 학생평가 활동에 대한 설계 및 성찰을 가능하게 하며 학생들의 특성에 맞는 개별적 평가를 하는 데에도 유용한 것으로 나타났다.

둘째, 이 연구의 평가척도는 교사의 학생평가 전문성의 인지적 요소 뿐만 아니라 평가에 대한 교사의 전반적인 태도 및 인식을 포함하여 보다 다각적인 측면에서 학생평가 전문성을 평가하기 위해 개발된 것이다. 이 연구는 개별 교사들의 학생평가 전문성을 보다 타당하고 객관적으로 평가할 수 있는 기준을 마련했다는 점에서 의의를 지닌다.

셋째, 교사의 학생평가 전문성 평가와 관련된 기존의 선행 연구에서

는 대부분 자기 보고식 설문 조사를 통한 ‘자기 평가 방법’을 활용한 반면, 이 연구에서는 학생평가 포트폴리오를 활용한 ‘전문가 평가 방법’을 통해 평가의 타당도와 신뢰도를 높인 점에서 의의가 있다. 전문가 평가 방법의 효과적인 활용을 위해서는 전문가 선정 및 채점자 훈련에 더욱 많은 관심과 세심한 주의가 요구되며 교사들에게는 전문가 평가 결과를 수용하려는 보다 개방적인 자세가 필요할 것이다.

넷째, 개발된 평가도구는 개별 교사의 학생평가 전문성에 대한 객관적인 평가 자료로 활용할 수 있을 뿐만 아니라 교사 스스로 자신의 학생평가 역량을 비교하고 진단하는데 유용할 것이다. 이는 효과적인 교원평가 및 교사 연수의 자료로도 활용할 수 있음을 시사한다.

한편, 이 연구에서 발견된 제한점 및 후속 연구를 위한 제언은 다음과 같다.

첫째, 이 연구는 서울, 경기, 인천 지역의 중·고등학교 영어과 교사만을 대상으로 교사의 학생평가 전문성을 평가할 수 있는 도구를 개발하고 타당화한 것으로 도출된 연구 결과를 다른 학교급 또는 다른 교과와 교사 전체로 일반화시키기 어렵다는 점이 한계점으로 지적된다. 최종적으로 연구에 참여한 교사 수는 36명으로, 한 학기 동안 광범위하고 다양한 자료를 수집하여 학생평가 포트폴리오를 구성하는데 있어서 많은 교사들이 심리적인 부담감을 느끼거나 많은 행정 업무로 인한 현실적인 어려움으로 연구 대상자 수를 늘리는데 제한점이 있었다. 또한, 같은 학교에서 근무하는 교사의 경우 공유하는 자료가 많아 교사별 학생평가 전문성의 차이가 제대로 드러나지 않는 등 표집의 대표성에 문제점이 발견되기도 하였다. 따라서, 더욱 객관적이고 신뢰로운 분석 결과를 도출하기 위해 후속연구에서는 다양한 학교에서 근무하는 보다 많은 수의 교사

들을 대상으로 평가도구의 타당도, 신뢰도 및 적합도를 검증할 필요가 있을 것이다.

둘째, 평가도구의 타당도와 신뢰도를 검증하는 과정에서 일부 한계점이 발견되었다. 먼저, 5명의 영어교육 전문가의 의견을 바탕으로 평가도구의 내용 타당도 검증이 이루어 진점을 들 수 있다. 보다 객관적인 검증을 위해 교육 측정·평가 전문가, 교육 현장의 경험이 많은 교육 관리자, 장학사 등을 포함하여 보다 많은 수의 다양한 전문가로부터 평가도구의 타당성에 관한 검토를 받을 필요가 있다. 다음으로, 평가도구의 문항 내적 신뢰도를 살펴보면 다른 하위 영역의 Cronbach 알파계수가 모두 .6 이상으로 양호한 반면 ‘평가 실행’ 영역의 경우 .56으로 비교적 낮은 편이었다. 이러한 연구 결과는 학생평가 전문성의 구성요소나 세부평가 기준의 수에 대한 보다 엄밀한 분석이 필요함을 시사한다.

셋째, 평가도구의 적합도 측면에서 살펴보면, 일부 평가 문항의 경우, 이론적으로 교사의 학생평가 전문성 평가 기준에 꼭 포함되어야 하는 항목이나 실제 학교 현장에서 활용하는 교사가 많지 않거나 학생평가 포트폴리오 자료에 제대로 반영되지 않은 경우가 존재하였다. 또한, 수집된 학생평가 포트폴리오만으로 교사의 학생평가 전문성을 다각도로 평가해야 하는 점에서 평가하고자 하는 교사의 실제 평가 전문성뿐만 아니라 보이지 않는 교육과정 또는 포트폴리오 자료 수집 및 정리에 있어서 교사의 성실성이나 꼼꼼함 등 학생평가 전문성 외 다른 특성이 전문가 평정 결과에 영향을 미치는 경우도 존재하였다. 특히, 학생평가 포트폴리오의 구성요소에 공정한 학생평가를 위한 노력 또는 평가 활동 전반에 관한 반성적 태도 등에 관한 교사 설문조사를 포함시켜 교사의 평가에 대한 태도 및 인식을 측정하고자 하였으나 주어진 자료만을 토대로 눈에 보이지 않는 교사의 심리적 특성을 전문가가 객관적으로 평가하기에는

어려움이 존재하였다. 후속 연구에서는 교사 자기평가 및 동료평가의 결과를 반영하고 심층 인터뷰나 수업 관찰 방법 등의 질적 연구를 병행하여 정량적, 정성적 측면에서 보다 체계적인 평가가 요구된다.

넷째, 이 연구에서는 중등학교 영어교사의 학생평가 전문성 평가 도구를 통해 전문가 평가자가 개별 교사들의 학생평가 전문성을 객관적으로 평가할 수 있는 있다는 점에서 의의가 있으나 교사 스스로 자신의 학생평가 전문성을 진단하기에는 평가도구의 복잡성이 지적되었다. 따라서, 학교 현장에서 교사들이 스스로의 학생평가 전문성을 비교적 효율적으로 평가하기 위한 보다 간소화된 형태의 평가도구 개발이 필요할 것이다. 평가도구의 신뢰도 검증을 위한 일반화가능도 이론의 D 연구 분석 결과를 바탕으로 할 때 비록 단일 차원성(uni-dimension)을 확인하지는 못했지만 평가 문항이 15개인 경우에도 적절한 수준의 신뢰도 계수를 확보하는 것으로 나타났다. 이러한 연구 결과는 각 영역 당 3문항으로 구성된 간편형 평가척도 개발 시에도 적절한 수준의 신뢰도를 확보할 수 있음을 시사한다.

마지막으로, 교사의 학생평가 전문성을 신장시키기 위한 구체적인 방안을 제시하면 다음과 같다. 첫째, 교사 차원의 노력으로는 학생평가 전문성의 하위 영역별로 자신의 강점과 약점에 관한 객관적인 파악을 통해 상대적으로 점수가 낮은 영역에 대해 보다 세심한 관심을 기울이고 평가 전문성을 개발하기 위한 다양한 활동에 적극적으로 참여할 필요가 있다.

둘째, 학교 차원에서는 동료평가나 전문가 평가방법을 활용하여 교사의 학생평가 전문성을 객관적으로 평가하고 교사 협의회, 평가 관련 교사 동아리나 전문적 학습 공동체의 활성화를 통해 교사의 학생평가 활동에 관한 피드백을 제공할 수 있는 기회를 확대할 필요가 있다.

셋째, 지역 교육청, 나아가 국가 차원의 노력을 살펴보면 다음과 같

다. 기존의 평가 관련 교사 연수는 교사의 개인 역량 강화에만 집중하여 동료교사나 학교 조직과의 연계가 전혀 이루어지지 않았다. 연수 이후 교사에게 피드백을 전혀 제공하지 않고 학교 현장에서 적용의 책임을 모두 교사에게 전가하는 등 일회성에 그치는 문제점이 지적되었다(김진규, 2009). 또한, 교사의 실천적 학생평가 전문성 신장에 큰 도움이 되지 않거나, 교사 연수의 커리큘럼, 프로그램 및 방법이 학교 현장에서 교사들의 학생평가 활동에서 실제로 강조되는 내용과 일치하지 않는 경우가 많았다. 따라서, 우선적으로 교사의 학생평가 전문성 신장을 위한 교사 커뮤니티를 활성화하기 위한 지원책을 마련해야 하며 현직 교사 연수(in-service teacher training) 뿐만 아니라 교사가 되기 위한 준비 과정인 예비 교사 연수(pre-service teacher training)에서도 교사의 평가 전문성을 신장시킬 수 있는 다양한 교사 교육 프로그램 및 정책이 마련되어야 할 것이다.

구체적인 방안으로 교생실습 내 교육과정(curriculum)에 학생평가 포트폴리오를 활용한 예비 교사 평가를 포함시키거나 교직과목이나 교과 교육학에서 다루는 일반적인 교육평가 과목 외에 학생평가 활동에서 필요한 소양과 기술을 보다 구체적으로 가르칠 수 있는 교과목을 개발하여 운영할 수 있을 것이다.

또한, 교사 개인적 특성 및 학교 특성에 따라 교사의 학생평가 전문성에 차이가 나타날 수 있음을 고려할 때, 교사들을 대상으로 한 요구조사를 바탕으로 특정 평가 영역에 초점을 두거나 교육 대상을 보다 구체화, 상세화한 교사 연수 프로그램을 시행한다면 보다 효과적인 성과를 가져올 것으로 기대된다. 아울러, 교사의 학생평가 전문성의 현재 역량을 파악하는데 그치지 않고 매학기 또는 매년마다 평가한 교사의 학생평가 전문성에 관한 데이터베이스를 구축하여 장기적인 관점에서 교사의

학생평가 전문성에 관한 관리 및 개발 노력이 요구된다.

참 고 문 헌

- 강양숙, 진승희, 장소영(2015). 중등영어교사의 서술형 평가에 관한 인식 및 운영 실태에 관한 연구. **영어교육연구**, 20(2), 323-353.
- 강애남, 이규민(2006). 학생들의 동료평가를 활용한 수행평가 결과의 일반화가능도 분석. **교육평가연구**, 19(3), 107-121.
- 교육부(2015). **2015 개정 교육과정 총론 및 각론 확정발표**. 교육부 보도 자료(2015.9.23.).
- 권순달(2010). 비교과 교사의 평가 전문성 분석. **교육방법연구**, 22(4), 219-236.
- 김경희, 김신영, 김성숙, 지은림, 반재천, 김수동(2006). 교사의 학생평가 전문성기준 개발. **교육평가연구**, 19(2), 89-112.
- 김경희, 박은아, 송미영, 상경아, 김수진, 김희경, 신진아, 서지영, 이채희, 김지영, 김기연, 최숙기(2012). 중등학교의 성취평가제 운영 방안 연구. 연구보고 CRE 2012-8. 서울: 한국교육과정평가원.
- 김기택, 조규희(2015). 영어교사의 전문성은 학생들의 영어 학습 성취도에 영향을 미치는가?. **초등영어교육**, 21(4), 63-83.
- 김길모, 김성수, 김성식(2014). 정보윤리 담당교사의 전문성 평가도구 개발. **한국정보교육학회**, 18(1), 1-11.
- 김동영(2006). 과학교사의 학생평가 전문성 신장을 위한 자기진단 프로그램 개발연구. 서울대학교 대학원 박사학위논문.
- 김선희(2006). 학생평가 전문성을 갖춘 수학교사 양성을 위한 수학학습평가 강좌의 교육 내용과 방법에 대한 제안. **학교수학**, 8(3), 301-326.
- 김선희(2012). 실습과 반성을 통한 수학 예비교사의 평가 전문성 신장. **수학교육학연구**, 22(2), 277-292.
- 김성숙, 김양분(2001). **일반화가능도이론**. 서울: 교육과학사.
- 김성숙, 김희경, 서민희, 성태제(2015). **교수·학습과 하나 되는 형성평가**.

서울: 학지사.

김성훈, 김신영, 서민원, 양길석(2009). 교육 프로그램 메타평가 기준 개발. **교육평가연구**, 22(3), 557-585.

김소현(2010). 포트폴리오 평가 문식성 진단을 위한 기준 개발-중등 읽기 영역 수업을 중심으로-, **교육연구**, 48, 5-25.

김수동, 이의갑, 김경희, 김선희, 박은아, 신명선, 김수진, 전영석, 박가나, 서수현(2005a). 교사의 학생평가 전문성 기준. 연구자료ORM 2005-51-1. 서울: 한국교육과정평가원.

김수동, 이의갑, 김경희, 김선희, 박은아, 신명선, 김수진, 전영석, 박가나, 서수현(2005b). 교사의 학생평가 전문성 신장 연구(Ⅱ). 연구보고 RRE 2005-3. 서울: 한국교육과정평가원.

김신영(2002). 현장교사의 평가전문성 연구. **교육평가연구**, 15(1), 67-85.

김신영(2004). **학생평가 전문성의 개념과 국가 수준의 기준 개발**. 교사의 학생평가 전문성 신장 모형 탐색 워크숍 자료. 서울: 한국교육과정평가원.

김신영(2006). 교원양성과정에서의 교육실습프로그램 평가 및 내실화 방안 연구. **교육논총**, 21, 5-22.

김신영(2007). 교사의 학생평가 전문성과 중등교사 양성과정. **교육평가연구**, 20(1), 1-16.

김신영, 송미영(2008). 예비교사의 학생평가 전문성과 교육실습의 효과, **교육평가연구**, 21(1), 79-97.

김신영(2010). 교육프로그램 메타평가를 위한 교육본위성 기준 개발, **교육원리연구**, 14(2), 한국교육원리학회.

김신영(2014). 학생평가에 대한 교사의 인식과 실천. **교육평가연구**, 27(1), 141-161.

김신영(2015). 교실내 학생평가의 운영실태와 개선 방향. **교육과정평가연구**, 18(3), 257-282.

김옥예(2006). 교사 전문성의 재개념화에 관한 연구. **교육행정학연구**,

24(4).139-160.

김은주(2002). EFL 상황에서의 바람직한 영어교사의 자질에 관한 연구.

영어교육연구, 14(1), 71-96.

김정환, 이계연(2005). 수업의 질 개선을 위한 교사 수업능력 자기평가
방략에 관한 논리적 고찰. **교육평가연구**, 18(3), 19-38.

김진규(2009). 학습공동체 활용 평가연수의 실천 전략, **교육평가연구**,
22(4), 939-959.

김진규(2013). 자유학기제를 위한 평가방법의 활용방안 탐색. **교육경북**,
155, 26-43.

김진석(2015). 영어로 진행되는 초등영어 수업 평가 요소에 관한 연구:
상호작용을 중심으로. **초등영어교육**, 21(4), 223-245.

김진완, 신명수, 어도선, 이소영, 이현우, 황종배(2004). 교원임용시험 표
준화방안: 중등영어과 임용시험. 2003 교과교육 공동연구지원 연
구결과 보고서.

김혜정(2008). 고등학교 국어과 평가의 문제점과 개선방안 -평가사례를
중심으로-. **국어교육학연구**, 32, 97-127.

김혜영(2004). 전자포트폴리오를 통한 교사 전문성 훈련 및 평가방안, **한
국교원교육연구**, 21(3), 113-134.

김혜영(2007). 영어교원양성 및 평가의 수행기준 연구. **외국어교육**, 14(4),
279-299.

박가나(2012). 학생평가에 대한 사회과 교사의 인식 및 평가 전문성 조
사. **사회과교육연구**, 19(1), 29-46.

박정(2013). 초등 예비교사의 학생평가 전문성 분석. **교육평가연구**
26(1), 15-33.

박지현, 조지민, 홍미영, 정혜경(2017). 학교수준 학생평가 운영 실태 및
인식 조사, **교육평가연구**, 30(1), 29-52.

박혜정, 최명숙, 이규민(2006). 가상강좌 강의평가 도구개발 및 타당화.
교육평가연구, 19(2), 203-223.

- 배호순(2008). **평가하는 학교사회: 한국교육평가체제를 말한다**. 서울: 교육과학사.
- 백순근(1999). 비디오 포트폴리오를 활용한 교수 능력 평가. **교육평가연구**, 12(2), 83-101.
- 백순근(2001).수행평가 정착 방안:방법 및 도구의 질 관리. **서울대학교 사대논총**, 62. 55-72.
- 백순근(2006). 학생평가에 대한 교사들의 학생부 신뢰도 제고 방안. 교육혁신위원회, 교사의 학생평가 전문성과 학생부 신뢰도 제고 방안 탐색. 한국교육평가학회 공동 개최 학술세미나 자료집. 61-80.
- 백순근, 임현수(2006). 중학교 교사용 실천지능검사의 개발 및 타당화. **교육평가연구**, 19(1), 223-240.
- 백순근, 함은혜, 이재열, 신호정, 유예림(2007). 중등학교 교사의 교수역량 구성요인에 대한 이론적 고찰. **아시아교육연구**, 8(1), 47-69.
- 백현정, 민찬규(2009). 중등학교 영어교사의 내용 교수 지식에 대한 조사 연구. **중등영어교육**, 2(2), 95-114.
- 소경희(2003). 교사 전문성의 재개념화 방향 탐색을 위한 기초연구. **교육과정연구**, 21(4),77-96.
- 손승남(2005). 교사의 수업전문성 관점에서 본 교사교육의 발전방향. **한국 교원교육연구**, 22(1), 89-108.
- 손영민(2007). 교사의 전문능력 개발을 위한 포트폴리오. **한국교원연구**, 24(2). 175-198.
- 손원숙, 신이나, 배주현, 박정(2015). 교사의 내외적 요인과 교실평가 간의 구조적 관계: 한국, 싱가포르 및 핀란드 비교, **교육평가연구**, 28(2), 601-622.
- 손유진, 황해익(2004). 디지털 포트폴리오 전략이 예비유아교사의 자기주도력 및 포트폴리오에 대한 인식에 미치는 영향. **열린유아교육연구**, 9(3), 29-48.

- 송미영, 김경희(2007). 교사의 학생평가 전문성 진단도구 개발. **교육과정 평가연구**, 10(1), 47-74.
- 송일섭, 김수란(2012). 다변량 일반화가능도 이론을 통한 인식론적 신념 검사 타당화 연구. **교육방법연구**, 24(1), 107-130.
- 서태열, 심광택, 이간용 (2005). 사회과 지리교사의 학생평가 전문성의 기준 설정에 관한 연구. **사회과교육**, 44(3), 127-150.
- 석은조(2006). 저널쓰기가 예비유아교사의 반성적 사고수준향상에 미치는 효과. **미래유아교육학회**, 13(4), 373-396.
- 성태제, 곽영순, 구영산, 권점례(2015). **2020 한국 초·중등교육의 향방과 과제: 교육과정, 교수·학습, 교육평가**. 서울:학지사.
- 신동일(2001). 일반화가능도 이론 적용을 중심으로 한 말하기 평가도구 타당도 검증 연구. **응용언어학**, 17(1), 199-221.
- 신종호, 조은별, 이유경(2012). 교육심리학에서의 목표 연구를 통해 본 교사의 역할과 전문성. **교육심리연구**, 26(1), 103-122.
- 안미화(2005). 초등학교 교사의 전문성 개발 척도. **한국교원교육연구**, 22(3), 281-306.
- 안진경(2011). 예비교사 교육과정에서 학습포트폴리오 활용의 효과. **열린 교육연구**, 19(4), 169-184.
- 양지승, 이규민(2007). 일반화가능도 이론을 적용한 단위검사 구성 검사 점수의 신뢰도 추정. **교육평가연구**, 20(1), 119-139.
- 어윤경(2006). 진로교육에 활용되는 자기보고식 심리검사의 faking-good 탐지기능분석. **교육평가연구**, 19(1), 241-263.
- 유진은(2012). 영어가 모국어가 아닌 학생용 영어쓰기 평가 연구: 교사 채점자 훈련과 채점자 일치도를 중심으로. **교원교육**, 한국교원대학교 교육연구원, 28(4), 63-80.
- 윤상이(2015). 초등교사의 학생평가 전문성이 학생의 수업 만족도에 미치는 영향. 서울대학교 대학원 석사학위논문.

- 윤승혜(2011). 비디오 포트폴리오 평가를 활용한 피드백 방식이 중등 교사의 실천적 교수역량에 미치는 영향. 서울대학교 대학원 석사학위논문.
- 이경아, 김경희(2006). 평생교육사 전문성 구인 타당화 및 전문성 형성에 영향을 미치는 요인 탐색에 관한 실증 연구. **평생교육학연구**, 12(2), 91-119.
- 이기훈(2014). **SPSS를 이용한 통계자료분석**. 서울: 자유아카데미.
- 이선영, 김성연, 김정하, 백근찬, 이병윤(2015). 다변량 일반화가능도 이론을 활용한 창의성 예비검사의 신뢰도 분석. **창의력교육연구**, 15(3), 83-107.
- 이병천, 김진석(2011). 영어로 진행되는 영어 수업의 평가 요소에 관한 연구. **초등영어교육**, 17(1), 145-170.
- 이영식(2000). 영어작문 평가에 대한 채점자 훈련의 원리. **영어교육**, 55(2), 35-56.
- 이인제, 이범홍, 박정, 진재관, 김옥남, 서수현, 김신영(2004). 교사의 학생 평가 전문성 신장 모형과 기준. 연구보고 RRE 2004-5-2. 서울: 한국교육과정평가원.
- 이재열(2008). 비디오 포트폴리오 평가가 중등 예비교사의 발표역량에 미치는 영향. 서울대학교 대학원 석사학위논문.
- 이종일(2003). 반성적 실행을 통한 사회과교사 자질개선. **사회과교육** 42(2), 5-27.
- 이희경(2007). 중등 영어교사들의 교수자로서의 자질과 평가자로서의 자질에 관한 연구. **영어교육**, 62(2), 309-330.
- 이혜정(2010). 공교육 내실화를 위한 교사 전문성 신장방안 탐색: 교사 견해를 중심으로. **열린교육연구**, 18(3), 77-102.
- 임수연(2014). 초등학교 영어교사의 전문성 평가도구 개발: 정의적 요인

- 을 중심으로. **영어교육**, 69(2), 221-246.
- 임승렬(2000). 교사교육에서 포트폴리오 개발과 활용. **한국교원교육**, 17(1), 315-335.
- 임찬빈, 이윤, 어도선(2006). **수업평가 매뉴얼: 영어과 수업평가 기준**. 서울: 한국교육과정평가원.
- 임찬빈, 장경숙, 이문복, 이해원(2011). 영어로 진행하는 초등영어 수업평가 도구 개발. **초등영어교육**, 17(1), 109-144.
- 임찬빈, 이해원, 주형미, 이영아, 이윤, 윤영애, 권영애(2013). 2009 개정 교육과정에 따른 초, 중학교 영어과 핵심 성취기준 개발 연구. 연구보고 CRC-2013-5. 서울: 한국교육과정평가원.
- 임찬빈, 이문복, 이해원, 김성혜, 주형미, 조보경, 김준식, 임수연, 우은정, 김병기, 장은경, 이영아, 김영현(2014). 2009 개정 교육과정에 따른 고등학교 영어과 핵심 성취기준 개발 연구. 연구보고 CRC 2014-5-4. 서울: 한국교육과정평가원.
- 장경숙(2008). 초등영어교사양성 과정에 관한 연구. **초등영어교육**, 14(1), 5-37.
- 조용개(2009). 대학 수업의 질적 개선 및 평가에 필요한 티칭 포트폴리오 구성요소 탐색. **교육공학연구**, 25(1), 1-27.
- 주세형(2011). 평가 문식성 신장을 위한 국어과 교사 교육. **문법교육**, 15, 31-50.
- 주순옥, 안경자(2015). 우수 초등학교 영어교사의 전문성 발달에 관한 사례 연구. **초등영어교육**, 21(2), 261-288.
- 정훈(2011). 교육 시장화 시대의 교사 전문성. **교육철학연구**, 33(3), 161-185.
- 전병만, 박준언, 안병규, 오준임, 유제명, 이소영, 김신혜(2005). 영어과 교사의 학생평가 전문성 신장을 위한 연구. **영어교육**, 60(2), 3-26.

- 전병만, 오준일(2006). 중등학교 영어과 교사의 학생평가 전문성에 대한 인식 조사. **영어교육연구**, 11(2), 19-43.
- 전병만, 장경숙, 박추원(2006). 웹기반 교수 포트폴리오를 활용한 영어 교사 평가. **한국외국어교육**, 13(4), 167-191.
- 지은림, 양명희, 정운선(2011). 교사의 수업 및 평가활동이 학생들의 자기 조절학습과 학업성취도에 미치는 영향. **초등교육연구**, 24(4), 165-184.
- 진경애, 이상하, 조보경(2009). TEE 진단평가 도구 개발. **영어교육연구**, 21(4), 315-334.
- 진경애, 김성혜, 이영아, 이정원, 권영애, 권희주, 강혜정, 조지형, 조정민(2016). 개정 교육교육 과정에 따른 초·중학교 영어과 평가 기준 개발 연구. 연구보고 CRC 2016-2-12. 서울: 한국교육과정평가원.
- 최금진(2012). 수업의 질 향상을 위한 티칭 포트폴리오. **공학교육**, 19(1), 39-43.
- 최윤영, 이경량, 김수연(2012). 예비 영어교사의 영어 수업 진행 능력 평가를 위한 평가 도구 개발 연구: 신뢰도·타당도·문항의 질 검증. **외국어교육**, 19(3), 315-343.
- 최지선(2014). 현직 수학교사 문항 개발 연수의 평가도구 개발 전문성 향상 효과. **수학교육학 연구**, 24(2), 253-267.
- 최진령, 이연승(2010). 유아교사와 초등교사의 교직 전문성에 관한 인식. **열린유아교육**, 15(1), 335-360.
- 하명애, 민찬규(2008). EFL 상황에서의 영어교사의 전문성 고찰. **영어교육교육**, 7(2), 87-110.
- 함은혜(2007). 교육실습의 교육적 가치에 대한 평가 연구. 서울대학교 대학원 석사학위논문.
- 허은정(2011). 학습조직으로서 학교가 교사전문성에 미치는 효과. **한국교**

원교육연구, 28(3), 29-53.

황정규(1998). **학교학습과 교육평가**. 서울: 교육과학사.

황은희(2008). 중등교사의 교수역량에 대한 경험적 연구. 서울대학교 대학원 박사학위논문.

황은희, 백순근(2008). 중등교사의 실천적 교수역량에 대한 자기평가와 전문가평가의 비교연구. **교육평가연구**, 21(2), 53-74.

American Federation of Teachers, National Council on Measurement in Education. National Education Association(AFT,NCME,NEA)(1990). Standard for teacher competence in educational assessment of students. **Educational Measurement: Issues and practice**, 9(4). 30-32.

Brennan, R. L.(1983). Manual for GENOVA: A GENeralized Analysis Of VAriance System version 2.1. Iowa City, IA: ACT, Inc.

Brennan, R. L.(2001). **Generalizability theory**. New York: Springer.

Brookhart, S. M. (2007). **Expanding views about formative classroom assessment: A review of the literature**. In J.H. McMilan(Ed)., **Formative classroom assessment: Theory into practice**. New York: Teachers College Press.

Brookhart, S. M. (2011). Educational assessment knowledge and skills for teachers. **Educational measurement: Issues and Practice**, 30(1), 3-12.

Brown, G. (2004). Teachers' conceptions of assessment: implications for policy and professional development. **Assessment in Education**, 11(3), 301-318.

Campbell, C., Murphy, J.A., & Holt, J.K.(2002). Psychometric analysis of an assesment literacy instrument: Applicability to pre-service teachers. Paper presented at the Annual Meeting of the

- Mid-Western Educational Research Association, Columbus, OH.
- Daniel, L.G., & King, D.A. (1998). Knowledge and use of testing and measurement literacy of elementary and secondary teachers. **The Journal of Educational Research**, 91(6), 331-344.
- Davison, C., & Leung, C. (2009). Current Issues in English Language Teacher-Based Assessment. **Tesol Quarterly**, 43(3), 395-415.
- DeLuca, C., & Klinger, D. A. (2010). Assessment literacy development: identifying gaps in teacher candidates' learning. *Assessment in Education: Principles. Policy & Practice*, 17(4), 419-438.
- DeLuca, C., Lapointe-McEwan, D., & Luhanga, U. (2016). Teacher assessment literacy: a review of international standards and measures. **Educational Assessment, Evaluation, and Accountability**, 28, 251-272.
- Freeman, D. (1989). Teacher training, development, and decision making: A model of teaching and related strategies for language teacher education. **TESOL Quarterly**, 23(1), 27-45.
- McMillan, J.H. (2014). **Classroom Assessment: Principles and Practice for Effective Standards-Based Instruction**. 6th Edition. Boston: Pearson.
- McMunn, N., McColskey, W., & Butler, S. (2004). Building teacher capacity in classroom assessment to improve student learning. **International Journal of Educational Policy, Research & Practice**, 4(4), 25-48.
- Mertler, C.A. (2003). Pre-service versus in-service teachers' assessment literacy: does classroom experience make a difference?. Paper presented at the Annual Meeting of the Mid-Western Educational Research Association (Columbus, OH, October 15-18, 2003).
- Mertler, C.A. (2005a), Secondary Teachers' Assessment Literacy: Does

- Classroom Experience Make a Difference?. **American Secondary Education**, 33(2),76-92.
- Mertler,C.A(2005b). The Role of Classroom Experience in Pre-service and In-service Teachers' Assessment Literacy. **Mid-Western Educational Researcher**, 18(4). 25-33.
- Mertler,C.A., & Campbell,C.(2005). Measuring teachers' knowledge and application of classroom assessment concept: development of the assessment literacy inventory. Paper presented at the annual meeting of the American Educational Research Association (Montréal, Quebec, Canada April 11-15, 2005).
- Jarr, K.A.(2012). Education practitioners' interpretation and use of assessment results. University of Iowa. Retrieved from <http://ir.uiowa.edu/etd/3317>.
- Joint Committe for Standards on Educational Evaluation.(2015). **Classroom assessment standards: practices for PK-12 teachers**. Kindle book.
- Kiomrs, R., Abdolmehdi.R., Naser, R.(2011). On the Interaction of Test Washback and Teacher Assessment Literacy: The Case of Iranian EFL Secondary School Teachers. **English Language Teacher**, 4(1), 156-161.
- Koh,K.H.(2011). Improving teachers' assessment literacy through professional development, **Teaching Education**, 22(3), 255-276.
- Plake,B.S.(1993). Teacher assessment literacy: Teachers' competencies in the educational assessment students. **Mid-Western Educational Research**, 6(1), 21-27.
- Plake.B.S., Impara,J.C. & Fager.J.J.(1993). Assessment Competencies of Teachers: A National Survey. **Educational Measurement: Issues and Practice**, 12(4), 10-39.

- Popham, W.J.(2011). Assessment literacy overlooked: A teacher educator's confession. **The Teacher Education**, **46**, 265-273.
- Richards, J.C.(2010). Competence and Performance in Language Teaching. **RELC Journal**, **41**(2), 101-122.
- Shin, S.(2012). "It Cannot Be Done Alone" : The Socialization of Novice Teachers in South Korea. **TESOL Quarterly**, **46**(3), 542-567.
- Shohamy, E. (2001). **The power of tests: A critical perspective on the uses of language tests**. London: Pearson.
- Shulman, L.S.(1987). Knowledge and teaching: Foundations of the new reforms. **Harvard Educational Review**, **57**(1). 1-22.
- Shulman, L.S.(1998). **Teacher portfolios: A theoretical activity**. In N. Lyons(ed.) *With Portfolio in Hand*, 23-27. New York: Teacher College Press.
- Stanley, p., & Murray, N.(2013). 'Qualified?' , A framework for comparing ELT teacher preparation courses. **Australian Review of Applied Linguistics**, **36**(1), 102-115.
- Stiggins, R.J.(1991). Assessment literacy. **Phi Delta Kappan**, **72**, 534-539.
- Stiggins, R.J (1999). Are you assessment literate?. **The High School Journal**, **6**(5), 20-23.
- Stiggins, R.J(2005). From formative assessment to assessment for learning: A path to success in standards-based schools. **Phi Delta Kappan**, **87**(4), 34-328.
- Swailers, S.(2003). "Professionalism: Evolution and Measurement". **The Service Industries Journal**, **23**(2). 130-149.
- Wallace, M.J.(1991). **Training foreign language teacher: A reflective approach**. Cambridge: Cambridge University Press.
- Webb, D.C.(2012). Teacher change in classroom assessment: The role of

teacher content knowledge in the design and use of productive classroom assessment. 12th International Congress on Mathematical Education, COEX, Seoul, Korea. retrieved March, 10, 2016 from the World Wide Web: <http://www.icme12.org/upload/UpFile2/TSG/1217.pdf>.

William, D., & Thompson, M. (2008). **Integrating assessment with learning: What will it take to make it work?**. In C.A. Dwyer (Ed.), *The future of assessment: Shaping teaching and learning* (pp. 53-82). New York: Lawrence Erlbaum Associates.

Xu, Y., & Brown, G. (2016). Teacher assessment literacy in practice: A reconceptualization. **Teaching and Teacher Education**, 58, 149-162.

Zhang, Z., & Burry-stock, J. A. (1997). Assessment practices inventory: a multivariate analysis of teachers' perceived assessment competency. Paper presented at the Annual Meeting of the National Council on Measurement in Education (Chicago, IL, March 25-27, 1997).

부 록 목 차

<부록 1> 교사 배경변수 설문지	135
<부록 2> 학생평가 포트폴리오 평가 설문지 (연구 참여 교사용)	136
<부록 3> 학생평가 포트폴리오 평가 설문지(전문가 평정자용)	137
<부록 4> 학생평가 전문성 평가척도(1차 예비 문항)	138
<부록 5> 학생평가 전문성 평가척도(2차 예비 문항)	144
<부록 6> 학생평가 포트폴리오 실제 자료	149
<부록 7> 교사의 학생평가 전문성 채점기준 (전문가 평정자용)	162
<부록 8> 교사의 학생평가 전문성 점수 분포(히스토그램)	175

[교사용 설문지]

<부록 1> 교사 배경변수 설문지

교사 ID: _____

다음은 학생평가 전문성에 영향을 미치는 변수를 알아보기 위해 선생님과 근무하시는 학교에 관한 간단한 설문 문항입니다. 해당하는 곳에 ✓표시를 하거나, 빈 칸에 내용을 적어주세요.

문항번호	
1	성별 : ① 남 ② 여
2	교직경력: ()년 ※예시: 10년 3개월인 경우 10.3으로 표기
3	학력: ① 학사 졸업 ② 석사 수료 ③ 석사 졸업 ④ 박사 수료 ⑤ 박사 졸업
4	전공: ① 영어 교육 ② 영문학 ③ 기타 전공: _____
5	평균 학급당 학생수: ()명
6	학교유형: ① 중학교 ② 일반계고 ③ 특성화고 ④ 특목고 ⑤ 기타: _____ ※예시: 자율형 사립고, 자율형 공립고 등
7	혁신학교 여부: ① 예 ② 아니오
8	수준별 수업 진행 여부: ① 예 ② 아니오
8-1	만약, 수준별 수업을 진행하신다면, 선생님께서 담당하시는 학급의 수준은? ① 상 ② 중 ③ 하

응답해 주셔서 감사합니다.

<부록 2> 학생평가 포트폴리오 평가 설문지 (연구 참여 교사용)

교사명: _____

소속학교: _____

연구에 참여해주셔서 다시 한번 감사드립니다. 다음은 선생님의 연구 참여에 관한 간단한 소감 및 향후 더 나은 학생평가포트폴리오 평가를 위한 설문입니다. 실제 경험을 바탕으로 선생님의 의견을 솔직하고 자세하게 답변해주시기 바랍니다.

1. 선생님께서 연구에 참여하시면서 가장 힘들었던 점은 무엇입니까?



2. ‘학생평가 포트폴리오’의 데이터 수집상 가장 어려웠던 항목은 무엇입니까?



3. 이번 연구에 참여하시면서 배운점이나 느낀 점은 무엇입니까? (학생 평가 포트폴리오를 통해서 배운 점은 무엇입니까?)



4. 학생평가 활동에서 가장 중요하다고 생각하시는 부분은 무엇입니까?
(평가계획, 평가도구 개발, 평가 수행, 평가 결과 분석 및 활용, 평가에 대한 태도 및 인식 영역 측면에서 구체적으로 작성해주시길 바랍니다.)



5. 선생님께서 기존에 참여하신 평가 관련 교사 연수(교내·외 현직연수, 워크숍, 온라인 연수 등)가 선생님의 학생평가 전문성 신장에 도움이 되었다고 생각하십니까? 만약 그렇지 않다면 평가 관련 교사연수에서 부족하거나 아쉬웠던 점은 무엇입니까?



응답해 주셔서 감사합니다 ☺

<부록 3> 학생평가 포트폴리오 평가 설문지(전문가 평정자용)

성명: _____

연구에 참여해주셔서 다시 한번 감사드립니다. 다음은 향후 더 나은 학생평가 포트폴리오 평가를 위한 설문입니다. 지난 평정 경험을 바탕으로 선생님의 의견을 솔직하고 자세하게 답변해주시기 바랍니다.

1. 선생님께서 평정하시면서(연구에 참여하시면) 가장 힘들었던 점은 무엇입니까?



2. ‘교사의 학생평가 전문성 평가척도’에서 불필요하다고 생각되시거나 평정이 어려웠던 평가 항목은 무엇이며, 그 이유는 무엇입니까?



3. 교사의 학생평가 포트폴리오를 활용한 교사의 학생평가 전문성 평가에 있어서 보완되어야 할 점이 있다면 무엇입니까? (평가의 타당도, 신뢰도, 현실적합성 등의 측면에서 구체적으로 서술해 주시길 바랍니다.)



응답해 주셔서 감사합니다 ☺

<부록 4> 학생평가 전문성 평가척도(1차 예비 문항)

구성요인	문항	포트폴리오자료	전혀 그렇지 않다	그렇지 않다	보통이다	그렇다	매우 그렇다
평가계획	1.[교육과정] 교사는 2009 개정 교육과정에서 제시하고 있는 영어과 핵심 성취 기준을 제대로 이해하고 있는가?	(A)(B) (C)(G) (L)	①	②	③	④	⑤
	2.[교육과정]교사는 영어과 교육과정을 평가 목표에 제대로 반영하고 있는가?	(A)(B) (C)(G) (L)	①	②	③	④	⑤
	3.[평가목표] 교사는 학습목표와 일치하는 평가 목표를 분명히 제시하고 있는가?	(A)(D) (E)	①	②	③	④	⑤
	4.[평가방법]교사는 학습목표와 일치하는 평가 방법을 선택할 수 있는가?	(A)(B) (C)	①	②	③	④	⑤
	5.[평가방법] 교사는 학생들의 수준에 적합한 평가 방법을 선택할 수 있는가?	(A)(B) (C)	①	②	③	④	⑤
	6.[평가의 실제성] 교사는 다양한 평가 방법의 이론을 바탕으로 실행 가능한 평가 과업 및 평가 문항을 제시하고 있는가?	(A)(B) (C)(I)	①	②	③	④	⑤
	7.[학습자의 인지적특성] 교사는 영어 학습자의 다양한 인지적 특성 및 관련 학습 이론을 제대로 파악하고 있는가?	(A)(G) (H)(I) (L)(M)	①	②	③	④	⑤
	8.[학습자의 정의적특성] 교사는 영어 학습자의 다양한 정의적 특성을 이해하고 평가 계획에 반영하는가?	(A)(B) (G)(H) (I)(L) (N)	①	②	③	④	⑤
	9.[통합적과업]교사는 교과서나 교수·학습과정에서의 과업 을 응용하여 평가활동을 계획하는가?	(B)(E) (G)	①	②	③	④	⑤
	10.[학습효과]교사는 각 평가방법이 학생들의 영어 학습에 미치는 긍정적인 영향과 부정적인 영향을 파악 하고 있는가?	(A)(B) (C)(D)	①	②	③	④	⑤

평 가 도 구 개 발	11.[평가계획서]교사는 체계적인 평가도구 개발을 위해 평가목적, 평가 영역, 내용, 평가도구 유형 등을 포함한 평가 계획서를 작성하는가?	(A)	①	②	③	④	⑤
	12.[도구제작계획]교사는 평가도구를 개발하기 위한 계획을 구체적으로 서술할 수 있는가?	(A)	①	②	③	④	⑤
	13.[평가도구이해]교사는 다양한 평가도구의 특징과 장·단점을 이해하고 있는가?	(A)(B) (C)(D) (G)(I) (L)(O)	①	②	③	④	⑤
	14.[평가과업]교사는 학생들의 성취수준에 적합한 평가 과업을 개발하여 제시하는가?	(B)(C) (G)	①	②	③	④	⑤
	15.[평가문항]교사는 학생들의 성취수준에 적합한 평가 문항을 개발하여 제시하는가?	(B)(C) (G)	①	②	③	④	⑤
	16.[채점기준]교사는 개발된 평가 과업과 평가 문항의 채점기준을 명확하게 설정하여 제시하는가?	(A)(D)	①	②	③	④	⑤
	17.[평가도구개발]교사는 각 언어 기능별로 핵심 성취 기준의 달성 여부를 점검하는 다양한 평가 도구를 직접 개발할 수 있는가?	(A)(D) (G)	①	②	③	④	⑤
	18.[평가도구선택]교사는 평가 목적에 따라 기존에 개발된 다양한 유형의 표준화 검사를 선택하고 활용할 수 있는가?	(A)(B) (C)(D) (E)	①	②	③	④	⑤
	19.[평가도구관리]교사는 평가도구의 타당도의 개념을 이해하고 확인하여 차후 평가도구 개발 자료로 활용 할 수 있는가?	(A)(I)	①	②	③	④	⑤
	20.[평가도구관리]교사는 평가도구의 신뢰도의 개념을 이해하고 확인하여 차후 평가도구 개발 자료로 활용 할 수 있는가?	(A)(I)	①	②	③	④	⑤
	21.[평가시기] 교사는 평가의 교육적 목적을 극대화시키기 위해 적절한 시기에 평가를 실시하는가?	(A)(B) (C)(L)	①	②	③	④	⑤
	22.[수업목표점검]교사는 수업 중 또는 수업 마지막 단계에서 학생들의 수업 목표 달성 정도를 평가하는가?	(A)(O)	①	②	③	④	⑤

평 가 실 행	23.[과제제시]교사는 수업 내용을 연습하거나 심화할 수 있는 과제를 제시하는가?	(A)G (L)M	①	②	③	④	⑤
	24.[진단평가]교사는 학생들의 선행 학습 수준을 미리 파악하기 위하여 진단평가를 실시하는가?	(A)G (L)M	①	②	③	④	⑤
	25.[형성평가]교사는 학생들의 성취수준에 적합한 형성평가를 실시하는가?	(A)G (L)M	①	②	③	④	⑤
	26.[비형식적평가]교사는 관찰, 면담, 학습 태도 체크리스트, 학습 유형 점검표, 학습일지 등의 비형식적 방법을 통해 학습자의 특성, 행동 및 학습목표 달성 여부에 관한 정보를 수집하는가?	(E)G (L)	①	②	③	④	⑤
	27.[학생안내]교사는 학생들에게 평가 실시 전에 미리 평가계획에 대해 자세하게 안내 하는가?	(E)	①	②	③	④	⑤
	28.[평가방법]교사는 중·고등학교 영어과 핵심 성취기준에 따라 각 성취수준별에 적합한 평가 방법을 이해하고 실시하는가?	(A)B (C)G (L)	①	②	③	④	⑤
	29.[평가방법]교사는 학생들의 흥미를 이끌 수 있는 다양한 평가 방법을 선택하여 실시하는가?	(G)L	①	②	③	④	⑤
	30.[자기평가]교사는 학생들에게 자기평가의 기회를 제공하는가?	(B)G (L)	①	②	③	④	⑤
	31.[동료평가]교사는 학생들에게 동료평가의 기회를 제공하는가?	(B)G (L)	①	②	③	④	⑤
	32.[물리적환경]교사는 평가의 공정성을 확보하기 위해 학생들에게 동일한 물리적 환경을 제공하는가?	(A)H	①	②	③	④	⑤
	33.[심리적환경]교사는 학생들의 개별적인 정의적 특성을 파악하여 특정 학생들이 평가에 대한 불안감이나 부담감에서 벗어 날 수 있도록 도움을 주는가?	(A)H	①	②	③	④	⑤
	34.[채점 훈련]교사는 서술형 평가의 유사답안 인정 및 부분점수 부여를 위한 채점기준을 일관성 있게 적용하기 위하여 사전에 동료 교사와 채점 연습 및 훈련을 실시하는가?	(F)①	①	②	③	④	⑤
	35.[채점의 정확성]교사는 서술형, 논술형 평가 및 수행평가의 채점기준에 근거하여 정확한 채점을 실시하고 있는가?	(D)J (K)	①	②	③	④	⑤

평 가 결 과	36.[채점]교사는 학습목표 및 평가목표에 따라 적절하게 채점 할 수 있는가?	Ⓐ Ⓓ ⓐ ⓓ ⓑ ⓔ ⓕ ⓓ	①	②	③	④	⑤
	37.[성적부여]교사는 평가에 따라 적합한 방식을 선택하여 점수를 부여할 수 있는가?	Ⓐ Ⓓ ⓐ ⓓ ⓑ ⓔ ⓕ ⓓ	①	②	③	④	⑤
	38.[결과분석]교사는 개발한 평가도구의 양적·질적 평가 결과에 관한 기초적인 통계자료를 분석할 수 있는가?	①	①	②	③	④	⑤
	39.[결과해석]교사는 평가 결과의 의미를 정확하고 객관적으로 해석 할 수 있는가?	ⓐ ⓓ ⓑ ⓔ ⓕ ⓓ	①	②	③	④	⑤
	40.[양적평가결과활용]교사는 수집된 양적 평가 결과를 활용하여 학생의 학습 성취 정도에 대한 종합적 판단을 내릴 수 있는가?	ⓐ ⓓ ⓑ ⓔ ⓕ ⓓ	①	②	③	④	⑤
	41.[질적평가결과활용]교사는 면담, 관찰 등을 통해 수집된 질적 평가 결과를 활용하여 학생의 학습 성취 정도에 대한 종합적 판단을 내릴 수 있는가?	ⓐ ⓓ ⓑ ⓔ ⓕ ⓓ	①	②	③	④	⑤
	42.[평가결과의 활용]교사는 수집된 양적·질적 평가 결과를 활용하여 차기 평가 계획 및 평가 활동에 적용하는가?	ⓐ ⓓ ⓑ ⓔ ⓕ ⓓ	①	②	③	④	⑤
	43.[표준화검사 결과해석]교사는 모의고사, 국가 수준의 성취도 평가나 수학능력 시험 등과 같은 표준화 검사의 결과를 정확하고 객관적으로 해석할 수 있는가?	ⓐ ⓓ ⓑ ⓔ ⓕ ⓓ	①	②	③	④	⑤
	44.[진단평가활용]교사는 학생의 학습 준비도와 현재 학습 상황을 점검하기 위해 진단평가를 활용하는가?	Ⓐ Ⓓ ⓐ ⓓ ⓑ ⓔ ⓕ ⓓ	①	②	③	④	⑤

분 석 및 활 용	45.[형성평가활용]교사는 형성평가를 실시하여 학생들의 영어 학습상의 어려움, 수업이해도, 오류 유형, 수업방법 및 수업 자료의 효율성 등을 파악하고 그 결과를 수업 활동 개선에 활용하는가?	GⓁ	①	②	③	④	⑤
	46.[총괄평가활용]교사는 수집된 총괄평가 결과를 활용하여 학생의 학습 성취 정도에 대한 종합적 판단을 내릴 수 있는가?	GⓁ MⓃ	①	②	③	④	⑤
	47.[학습개선]교사는 영어 학습과 관련 있는 학생의 인지적 특성 및 정의적 특성을 향상시키는 방향으로 평가 결과를 활용하는가?	LⓂ	①	②	③	④	⑤
	48.[동료교사와의 협의]교사는 동료교사와 평가 결과가 일치하지 않을 경우 그 원인을 분석하고 해석할 수 있는가?	Dⓕ ①	①	②	③	④	⑤
	49.[교수기술 및 전략]교사는 평가 결과를 바탕으로 수업의 내용, 교수 자료, 교수 방법 등을 조절하는가?	①ⓐ	①	②	③	④	⑤
	50.[수업계획]교사는 평가 결과를 바탕으로 학생들의 수준에 맞는 수업을 계획할 수 있는가?	①ⓐ	①	②	③	④	⑤
	51.[평가결과의 보고]교사는 평가 실시 후 학생 수준 또는 집단 수준(학급, 학교, 교육청, 국가수준)에서 평가 결과를 해석하고 설명할 수 있는가?	①Ⓜ N	①	②	③	④	⑤
	52.[과제피드백]교사는 수업시간에 과제에 관한 피드백을 적절하게 제공하는가?	GⓁ	①	②	③	④	⑤
	53.[학생피드백]교사는 평가결과에 기초하여 개별 학생들에게 학생의 강점과 약점에 대해 설명하고 개선 방향을 제시하는가?	GⓁ	①	②	③	④	⑤
	54.[형성평가피드백]교사는 평가 결과를 바탕으로 학습자의 질문이나 형성평가에 대한 피드백을 적절하게 제공하는가?	Gⓙ KⓁ MⓃ	①	②	③	④	⑤
	55.[학부모와의 의사소통]교사는 평가와 관련된 정보를 학부모와 공유하고 학생들의 학습 개선을 지원 할 수 있는 방법을 제안하는가?	EⓂ	①	②	③	④	⑤
	56.[평가결과의 교사간협의]교사는 평가 실시 후 동료 영어 교사와 평가절차나 결과에 대해 논의하는가?	①	①	②	③	④	⑤

평 가 에 대 한 태 도 및 인 식	57.[윤리적·법적책임]교사는 평가 문항 공개 및 학생들의 개인정보에 관한 철저한 보안을 포함한 학교·교육청·국가수준의 평가 관련 지침 및 법적 책임에 관해 명확히 이해하고 이를 학생평가 활동에 적용하는가?	Ⓐ(H) ①	①	②	③	④	⑤
	58.[평가공정성]교사는 학생 존중을 바탕으로 학생 평가 활동 시에 학생 성적에 직·간접적으로 영향을 줄 수 있는 행위를 최소화하는가?	Ⓐ(H) ㉔(O)	①	②	③	④	⑤
	59.[성적관리규정]교사는 평가활동 시 학교의 성적관리 규정을 준수하고 있는가?	Ⓐ(H) ㉔(O)	①	②	③	④	⑤
	60.[평가의공정성]교사는 평가결과에 영향을 미치지 않도록 평가도구, 학습 방법, 평가 기획 및 평가 방법을 공정하게 제공하는가?	Ⓐ(H) ㉔(O)	①	②	③	④	
	61.[평가활동에 대한 반성]교사는 자신의 전반적인 교실 내 평가 활동을 반성하고 발견된 문제점에 대한 대안을 모색하고 교수·학습 개선 방안을 스스로 제시 할 수 있는가?	㉔(O)	①	②	③	④	⑤
	62.[전문성 개발]교사는 학교 내 동료교사와의 협력을 통해 평가전문성을 개발하려는 노력을 기울이는가?	㉔(F)	①	②	③	④	⑤
	63.[전문성개발]교사는 평가 전문성 향상을 위해 외부의 평가 관련 워크숍, 세미나, 멘토링, 교과협의회 등에 적극적으로 참여하는가?	㉔(F)	①	②	③	④	⑤
	64.[동료교사와의 협업능력]교사는 평가 과정에서 교내·외 전문가 집단과 평가 절차, 채점기준, 평가 결과, 교수·학습 개선 방법 등에 대해 논의하고 협력하는가?	㉔(㉔) ①	①	②	③	④	⑤
	65.[평가의 환류적 기능]교사의 전반적인 학생평가 활동이 학생들의 학습에 긍정적인 영향을 미치는 방향으로 실시되었는가?	전체 자료	①	②	③	④	⑤

<부록 5> 학생평가 전문성 평가척도(2차 예비 문항)

구성요인	문항	포트폴리오자료	전혀 그렇지 않다	그렇지 않다	보통이다	그렇다	매우 그렇다
평가계획	1.[교육과정]교사는 2009 개정 교육과정에서 제시하고 있는 영어과 핵심 성취 기준을 제대로 이해하고 있으며, 교육과정에 반영할 수 있는 역량을 지니고 있는가?	(A)(B) (C)(G) (L)	①	②	③	④	⑤
	2.[평가목표]교사는 학습목표와 일치하는 평가 목표를 분명히 설정하고 이를 구체적으로 제시하고 있는가?	(A)(D) (E)	①	②	③	④	⑤
	3.[평가의 실제성]교사는 다양한 평가 방법의 이론을 바탕으로 실행 가능한 평가 과업 및 평가 문항을 제시하고 있는가?	(A)(B) (C)(I)	①	②	③	④	⑤
	4.[학습자의 인지적특성]교사는 영어 학습자의 다양한 인지적 특성 및 관련 학습 이론을 제대로 파악하고 있는가?	(A)(G) (H)(I) (L)(M)	①	②	③	④	⑤
	5.[학습자의 정의적특성]교사는 영어 학습자의 다양한 정의적 특성을 이해하고 평가 계획에 반영하는가?	(A)(B) (G)(H) (I)(L) (N)	①	②	③	④	⑤
	6.[통합적과업]교사는 교과서나 교수·학습과정에서의 과업을 응용하여 평가활동을 계획하는가?	(B)(E) (G)	①	②	③	④	⑤
평가도	7.[도구제작계획]교사는 체계적인 평가 도구 개발을 위해 평가목적, 평가 영역, 내용, 평가 도구 유형 등을 포함한 평가 계획서를 작성하는가?	(A)	①	②	③	④	⑤
	8.[평가도구이해]교사는 다양한 평가 도구의 특징과 장·단점을 이해하고 있는가?	(A)(B) (C)(D) (G)(I) (L)(O)	①	②	③	④	⑤
	9.[평가수행맥락]교사는 학생들의 성취수준에 적합한 평가과업 및 문항을 개발하여 제시하는가?	(B)(C) (G)	①	②	③	④	⑤

구 개 발	10.[채점기준]교사는 개발된 평가 과업과 평가 문항의 채점기준을 명확하게 설정하여 제시하는가?	(A)(D)	①	②	③	④	⑤
	11.[평가도구개발]교사는 각 언어 기능별로 핵심 성취 기준의 달성 여부를 점검하는 다양한 평가 도구를 직접 개발할 수 있는가?	(A)(D) (G)	①	②	③	④	⑤
	12.[평가도구선택]교사는 평가 목적에 따라 기존에 개발된 다양한 유형의 표준화 검사를 선택하고 활용할 수 있는가?	(A)(B) (C)(D) (E)	①	②	③	④	⑤
	13.[평가도구관리]교사는 평가 문항의 특성과 평가도구의 양호도의 개념을 이해하고 확인하여 차후 평가 도구 개발 자료로 활용 할 수 있는가?	(A)(I)	①	②	③	④	⑤
평 가 실 행	14.[평가시기]교사는 평가의 교육적 목적을 극대화시키기 위해 적절한 시기에 평가를 실시하는가?	(A)(B) (C)(L)	①	②	③	④	⑤
	15.[숙제 및 형성평가]교사는 수업 내용을 연습하거나 심화할 수 있는 과제를 제시하고 학생들의 성취수준에 적합한 형성평가를 실시하는가?	(A)(G) (L)(M)	①	②	③	④	⑤
	16.[비형식적평가]교사는 관찰, 면담, 학습 태도 체크리스트, 학습 유형 점검표, 학습일지 등의 비형식적 방법을 통해 학습자의 특성, 행동 및 학습목표 달성 여부에 관한 정보를 수집하는가?	(G)(L)	①	②	③	④	⑤
	17.[학생안내]교사는 학생들에게 평가 실시 전에 미리 평가 시기, 평가도구, 채점기준, 평가 실시방법 등에 대해 자세하게 안내 하는가?	(E)	①	②	③	④	⑤
	18.[평가방법]교사는 중·고등학교 영어과 핵심 성취기준에 따라 각 성취수준별로 적합한 다양한 평가방법을 이해하고 실시하는가?	(A)(B) (C)(G) (L)	①	②	③	④	⑤
	19.[평가참여]교사는 학생들에게 자기평가 및 동료 간 상호 평가의 기회를 제공하는가?	(B)(G) (L)	①	②	③	④	⑤
	20.[물리적환경]교사는 평가의 공정성을 확보하기 위해 학생들에게 동일한 물리적 환경을 제공하는가?	(A)(H)	①	②	③	④	⑤
	21.[심리적환경]교사는 학생들의 개별적인 정의적 특성을 파악하여 특정 학생들이 평가에 대한 불안감이나 부담감에서 벗어 날 수 있도록 도움을 주는가?	(A)(H)	①	②	③	④	⑤

평가 결과 분석 및 활용	22.[채점훈련]교사는 서술형 평가의 유사답안 인정 및 부분 점수 부여를 위한 채점기준을 일관성 있게 적용하기 위하여 사전에 동료 교사와 채점 연습 및 훈련을 실시하는가?	F①	①	②	③	④	⑤
	23.[채점의 정확성]교사는 서술형, 논술형 평가 및 수행평가의 채점기준에 근거하여 정확한 채점을 실시하고 있는가?	D② K	①	②	③	④	⑤
	24.[채점 및 성적부여]교사는 학습목표 및 평가 목표에 따라 적절한 채점 방식 및 점수 산출방식을 선택하여 실시하는가?	A③ G④ I⑤ K⑥	①	②	③	④	⑤
	25.[결과분석 및 해석]교사는 개발한 평가 도구의 양적·질적 평가 결과에 관한 기초적인 통계자료를 분석하고 그 의미를 정확하고 객관적으로 해석 할 수 있는가?	①	①	②	③	④	⑤
	26.[평가결과활용]교사는 수집된 양적·질적 평가 결과를 활용하여 학생의 학습 성취 정도에 대한 종합적 판단을 내릴 수 있으며 차기 평가 계획 및 평가 활동에 적용하는가?	G① L② N	①	②	③	④	⑤
	27.[표준화검사 결과해석]교사는 모의고사, 국가 수준의 성취도 평가나 수학능력 시험 등과 같은 표준화검사의 결과를 정확하고 객관적으로 해석 할 수 있는가?	I③ M	①	②	③	④	⑤
	28.[진단평가활용]교사는 학생의 학습 준비도와 현재 학습 상황을 점검하기 위해 진단평가를 활용 하는가?	A④ G	①	②	③	④	⑤
	29.[형성평가활용]교사는 형성평가를 실시하여 학생들의 영어 학습상의 어려움, 수업이해도, 오류 유형, 수업 방법 및 수업 자료의 효율성 등을 파악하고 그 결과를 수업 활동 개선에 활용 하는가?	G⑤ L	①	②	③	④	⑤
	30.[총괄평가활용]교사는 수집된 양적·질적 평가 결과를 활용하여 학생의 학습 성취 정도에 대한 종합적 판단을 내릴 수 있는가?	G⑥ M⑦ N	①	②	③	④	⑤
	31.[학습개선]교사는 영어 학습과 관련 있는 학생의 인지적 특성 및 정의적 특성을 향상시키는 방향으로 평가 결과를 활용하는가?	L⑧ M	①	②	③	④	⑤

	32.[교수기술 및 전략]교사는 평가 결과를 바탕으로 수업의 내용, 수준(보충/심화), 교수 자료, 교수 방법 등을 조절하는가?	①②	①	②	③	④	⑤
	33.[평가결과의 보고]교사는 평가 실시 후 학생 수준 또는 집단 수준(학급, 학교, 교육청, 국가수준)에서 평가 결과를 해석하고 설명할 수 있는가?	①② ③④ ⑤⑥	①	②	③	④	⑤
	34.[과제피드백]교사는 수업시간에 학생들에게 과제에 관한 피드백을 적절하게 제공하는가?	③④	①	②	③	④	⑤
	35.[학생피드백]교사는 평가 결과를 바탕으로 학습자의 질문이나 형성평가에 대한 피드백을 적절하게 제공하는가?	③④ ⑤⑥ ⑦⑧	①	②	③	④	⑤
	36.[학부모와의 의사소통]교사는 평가와 관련된 정보를 학부모와 공유하고 학생들의 학습 개선을 지원 할 수 있는 방법을 제안하는가?	⑤⑥	①	②	③	④	⑤
	37.[평가결과의 교사간협의]교사는 평가 실시 후 동료 영어 교사와 평가절차나 결과에 대해 논의하는가?	①	①	②	③	④	⑤
평가에 대한 태도 및 인식	38.[윤리적·법적책임]교사는 평가 문항 공개 및 학생들의 개인정보에 관한 철저한 보안을 포함한 학교의 성적 관리 규정, 교육청·국가수준의 평가 관련 지침 및 법적 책임에 관해 명확히 이해하고 이를 학생평가 활동에 적용하는가?	③④ ⑤⑥ ⑦⑧	①	②	③	④	⑤
	39.[평가공정성]교사는 학생 존중을 바탕으로 학생 평가 활동 시에 학생 성적에 직·간접적으로 영향을 줄 수 있는 행위를 최소화하며 개별 학생 특성을 고려하여 평가도구, 학습 방법, 평가 기회 및 평가 방법을 공정하게 제공하는가?	③④ ⑤⑥ ⑦⑧	①	②	③	④	⑤
	40.[평가활동에 대한 반성]교사는 자신의 전반적인 교실내 평가 활동을 반성하고 발견된 문제점에 대한 대안을 모색하고 교수·학습 개선 방안을 스스로 제시 할 수 있는가?	①②	①	②	③	④	⑤
	41.[전문적 자율성]교사는 자신의 학생평가 전문성을 개발하기 위한 활동에 활발하게 참여하는가?	⑦	①	②	③	④	⑤

42.[동료교사와의 협업능력]교사는 평가 과정에서 교내·외 전문가 집단과 평가 절차, 채점기준, 평가 결과, 교수·학습 개선 방법 등에 대해 논의하고 협력하는가?	<div> <div>①</div> <div>②</div> <div>③</div> <div>④</div> <div>⑤</div> </div> <div> <div>①</div> <div>②</div> <div>③</div> <div>④</div> <div>⑤</div> </div>
43.[평가의 환류적 기능]교사의 전반적인 학생평가 활동이 학생들의 전인적인 성장과 변화에 중점을 두고 긍정적 역류효과(positive washback)를 극대화시키는 방향으로 계획 및 실행되었는가?	<div> <div>①</div> <div>②</div> <div>③</div> <div>④</div> <div>⑤</div> </div> <div> <div>①</div> <div>②</div> <div>③</div> <div>④</div> <div>⑤</div> </div>

<부록 6> 학생평가 포트폴리오 실제 자료: ⑤과제 및 학습태도 체크리스트
예시

번호	성명	교과서단어시험				올림포스단어시험						학습지			참여도 (도장 판)	태도
		5과	6과	7과	8과	5강	6강	8강	9강	10강	총점	1차	2차	총		
1		20	19	19	20	20	20	19	18	20	174.5	15	16	31	14	
2		20	19	20	20	20	19	20	19	19	175.5	15	16	31	10	
3		20	19	20	20	20	19	20	20	20	178	17	17	34	23	
4		19	18	19	20	19	12	17	17	18	158.5	15.5	16	31		
5		18	19	17	17	19	20	17	20	18	165	14	17	31	5	
6		3	4	9	17	7	19	17	18	16	110	15.5	15	30.5	6	
7		8	4	9	16	4	1	17	18	16	93	13	15	28	9	
8		16	19	19	18	15	17	20	20	20	163.5	8	0	8	4	
9		20	19	19	19	20	18	19	16	18	168	15.5	14	29.5	13	
10		19	19	20	20	20	18	19	20	20	174.5	15	16	31	6	
11		0	9	0	4	3	5	8	3	2	34	8	6	14	3	주의산만 -3
12		20	19	19	19	17	20	17	20	19	170	16	15	31	11	
13		11	7	7	6	7	3	18	18	15	92	13	15	28	4	
14		14	19	17	3	10	14	18	14	17	125.5	16	15	31	5	참 -4
15		16	17	12	15	9	6	5	16	16	111.5	14	16	30	5	
16		20	19	20	20	20	20	19	20	20	178	16	14	30	5	
17		20	19	20	20	20	16	20	20	20	174.5	13	17	30	15	
18		10	17	11	11	12	11	4	6	14	96	9	14	23	1	
19		8	6	7	19	10	16	11	8	9	94		6.5	6.5	11	(축구부)
20		0	4	11	0	2	1	13	10	11	52	3	11	14	3	교과서-1, 교재-3
21		2.5	9	10	7	3	4	1	12	8	56.5		5.5	5.5	1	과제-2, 교재-2
22		16	19	20	20	20	20	20	20	20	175	15	17	32	25	
23		14	15	19	18	14	13	19	18	20	150	9	15	24	11	
24		20	19	20	20	20	20	20	20	20	179	14	17	30.5	13	
25		0	5	5	1	0	4	8	12	11	46	9	12	21	5	
26		4.5	9	3	5	4	4	13	15	13	70.5	8	13	20.5	13	
27		20	19	19	19	20	20	20	20	20	176.5	16	17	33	19	
28		3	5	2	2	5	3	6	4	6	36		12	12	3	(축구부)
29		11	18	19	17	5	10	17	19	18	134	15	17	32	14	과제-1
30		2	1	0	2	5	2	0	1	1	14		0	0	1	참-5, 교과서-3, 교재-2
32		17	17	19	16	8	10	18	18	20	143	14	12	26	14	
33		15	13	15	12	7	11	15	16	18	122	10	11	20.5	6	
34		19	19	20	20	20	20	20	20	20	178	13.5	17	30.5	21	

ॐ

박	박
김	이

안;	문
김	박

2

■ 개인별 과제 (5반)


■ 개인별 과제 (5반)


1. 프런트
2. 단서망 (3개씩)
3. 단어 (1-30)

<부록 6> 학생평가 포트폴리오 실제 자료: ⑧비공식적 학생평가 자료 및 기록물 예시

00반 비공식적 학생평가 자료 및 기록물(평소 비공식적 모둠별 단어시험 결과 기록 및 수업 발표 자료)

정O현 (멘토)	이O재	김O영 (멘토)	임O원	윤O현 (멘토)	강O범	나O지 (멘토)	김O지
박O선	여O아	권O주	민O현	정O훈	김O연	이주한	안정빈
<모둠별 활동> 수업 발표: 6과 모둠별단어시험결과: 1단원: 갑지1(별점) 2단원: 30.25 3단원: 38 (1등) 4과단어: 29.75(2등)		<모둠별 활동> 수업 발표: 7과 모둠별단어시험결과: 1단원: 16.25 2단원: 22 3단원: 27.5 4과단어 27.5(3등)		<모둠별 활동> 수업 발표: 5과 모둠별단어시험결과: 1단원: 19(3등) 2단원: 32.3(2등) 3단원: 33 4과단어 27		<모둠별 활동> 수업 발표: 4과 모둠별단어시험결과: 1단원: 16 2단원: 23 3단원: 26 4과단어 23	
유O은 (멘토)	길O찬	심O은 (멘토)	임O주	한O림 (멘토)	최O우	최O지 (멘토)	이O아
강O구	정O은	이O호	이O성	장O혁	김 O	박준수	이준성
<모둠별 활동> 수업 발표: 10과 모둠별단어시험결과: 1단원: 1등 2단원: 30.5(3등) 3단원: 34 (3등) 4과단어: 33.5(1등)		<모둠별 활동> 수업 발표: 9과 모둠별단어시험결과: 1단원: 15 2단원: 23 3단원: 27 4과단어: 24		<모둠별 활동> 수업 발표: 8과 모둠별단어시험결과: 1단원: 18.75 2단원: 34.3(1등) 3단원: 35.5 (2등) 4과단어: 26.5		<모둠별 활동> 수업 발표: 2과 모둠별단어시험결과: 1단원: 12.7 2단원: 20.75 (별칙) 3단원: 23.7 (별칙) 4과단어: 22.75(별칙)	
		서O정 (멘토)	박O은				
		정O름	이O민				
		최O석					
		<모둠별 활동> 수업 발표: 3과 모둠별단어시험결과: 1단원: 19.8(2등) 2단원: 24 3단원: 28.5 4과단어: 25					

주제	Seeds of change : planting a path to peace.	date	9 / 11	교사확인 
학습내용	<p>encourage A to B. → A에게 B하라고 격려했다.</p> <p>she also held seminars to <u>encourage</u> citizens to hold government officials accountable for ~</p> <p>She <u>started out</u> planting trees <u>and</u> found herself in the front line of fighting for the democratic movement in her country</p> <p>she started out ①] and 병2 found ②</p> <p>" <u>Given</u> this new climate, ~ ↳ 고려하다.</p>			
소감	<p>난 이제까지 Given이란 단어가 주가 가는 단어 뜻을 생각했는데 <u>함께</u> 사용된 Given이 고려하다 라는 뜻도 알게되니 <u>좋았다</u>.</p>			

주제	포스트 1강 자칭추론	date	8 / 22	교사확인 
학습내용	<ul style="list-style-type: none"> No other <u>activity</u> provides so many health benefits with so few dangers. (가장 적은 위험으로, 많은 건강상의 이점을 제공한다) have nothing to do : 관련이 없다 be bound to ~ : ~할 것이다. 반드시 ~해야 한다 ex) Our team is bound to win (우리팀이 분명히 이길 것이다) We are bound to find him (우리는 반드시 그를 찾을 것이다) enjoy doing (to. x) want to (ing. x) still to amuse : ~가 재미있을 정도로 			
소감	<p>나는 문법뿐만 아니라 단어도 모르는 단어가 많아서 비교적 비교적 내용이 쉬운 교재에 보다는 포스트를 더 어려워했었는데, 수업을 따라갈 수 있어 문법을 하문장 한문장씩 천천히 알려주셔서 새로운 단어들을 알 수 있어서 좋았다.</p>			

<부록 6> 학생평가 포트폴리오 실제 자료: ⑨수행평가 및 지필평가에
대한 학생 안내자료 예시

안녕하세요, 학생들! 생각보다 착한 선생님의 발표 과제 상제 공지입니다.
 체험학습으로 지친 학생들을 배려하여 제출 기간을 1주일 연장합니다. 짹짹!
 대신 수준 높은 발표와 뜨거운 열정을 기대합니다. 학생들의 희망대로 동료평가 없이 선생님
 혼자 평가합니다. 평가 요소 세부사항을 안내하니, 작업 하실 때 참고 하시기 바랍니다.
 그럼 다들 힘내세요 :) - 2016. 아름다운 가을. - 쌤

1. 제출 기간: 10.24(월) 오전 9시까지
2. thejinsook7@gmail.com으로 제출 (thejinsook@gmail.com 으로 받지 않습니다!!!)
3. 가) 메일제목: 30701 Presentation Video 제출
나) 파일명: 30701 Presentation Video

평가 요소 세부 사항		점수
과제 완성	3~4분 분량을 잘 준수하였는가?	-0.5
	10.24(월) 오전 9시, 기한을 잘 지켰는가?	-0.5
	thejinsook7@gmail.com 에 메일 제목 및 파일명을 정확히 기입하여 잘 보냈는가?	-0.5
언어	발음 및 발화 속도는 적당한가?	-0.5
	유창성 및 발표력은 적당한가?	-0.5
	표현이나 문법 요소는 적당히 사용되었는가?	-0.5
	표절 없이 자신의 언어로 잘 설명하였는가?	-0.5
내용	관심 있는 세계 문제가 포함되었는가?	-0.5
	세계문제를 해결하고자 하는 기관이나 프로젝트 소개는 적절한가?(who, what, how, result)	-0.5
	본인의 이야기가 적절하게 포함되었는가? (관심을 갖게 된 이유, 알게 된 이후의 변화 등등)	-0.5
발표 자료	발표 자료는 효과적으로 사용되었는가?	-0.5
	완전 창의적이거나 인상적인 발표 자료!	+0.5

★ 최고점 10점에서 위의 감점 요소를 반영한 후 반올림하여 성적에 반영합니다. 착하죠?
(ex. 감점 요소 1개인 경우 9.5 → 10점, 2개인 경우 9.0 → 9.0점.
3개인 경우 8.5 → 9.0점, 4개인 경우 8.0 → 8.0점)

친절한 선생님의 쓰기 과제 제출 방법 및 평가 요소 세부 사항 공지

1. 제출기간: 2nd draft 11.07(월) 8:50.a.m. (5반 11.06(일))
3rd draft 11.14(월) 8:50.a.m. (5반 11.13(일))
Final ver.11.21(월) 8:50.a.m. (5반 동영, 금요일 수업이 있는 3, 4반을 배려하여 제출 기한 연기!)
2. 제출방식: 가) thejinsook7@gmail.com으로 제출
나) 매일 제목: 30701 조진숙 Writing Project 제출
다) 파일명: 30701 조진숙 Writing Project
3. 제출사항(Final Version)
 - 가) 청원서 혹은 제안서 1부
 - 1) 분량: 1 to 1.5 page
 - 2) 양식: MS Word, 여백(4면) 2.54cm, Times New Roman, 12pt, 줄 간격 1.15
 - 3) 구성: 제목, 도입(자기소개 및 문제 제기), 주장(제안 사항), 근거 3, 결론, 받는 대상
 - 나) 온라인 청원사이트(change.org, ipetitions.com)에 제출 혹은 대상자에게 이메일 보낸 인증샷

평가 영역	평가 요소 세부 사항	점수
과제 완성 (2.5점)	최종 제출, 11.21(월) 오전 08:50 기한을 잘 지켰는가?	-0.5
	proposal, 1 st draft(아웃라인), 2 nd draft, 3 rd draft의 제출 기한을 잘 지켰는가? (2회 이상 기한 내 미제출시 감점)	-0.5
	매일 제목 및 파일명을 정확히 기입하여 thejinsook7@gmail.com에 보냈는가?	-0.5
	분량과 양식을 둘 다 잘 맞추었는가?	-0.5
	온라인 청원 사이트에 청원서를 올렸는가?	-0.5
	제안서를 받을 대상에게 이메일을 통해 보냈는가?	
내용 (4점)	주장하는 바를 잘 요약하는 제목을 사용하였는가?	-0.5
	문제제기와 자기소개가 들어간 도입이 잘 사용되었는가?	-0.5
	본인의 주장이 명확하게 드러났는가?	-0.5
	근거 1: 주장을 뒷받침하기에 충분한가?	-0.5
	근거 2: 주장을 뒷받침하기에 충분한가?	-0.5
	근거 3: 주장을 뒷받침하기에 충분한가?	-0.5
	본인의 주장과 근거를 결론에서 잘 정리하였는가?	-0.5
	적합한 대상에게 자신의 주장하는 바를 전달하고 있는가?	-0.5
언어 (2점)	문장을 구성하는 문법 요소가 적절히 사용되었는가?	-0.5
	어휘나 표현이 적절히 사용되었는가?	-0.5
	표절 없이 자신의 언어로 잘 설명하였는가?	-0.5
	APA style에 맞게 citation을 잘 표시하였는가?	-0.5
해결책	주장 하는바(세계 문제에 대한 해결책)에 실효성이 있는가?	-0.5
	효과적이고 참신한 해결책을 제안	+0.5

★★★★★ 위의 평가요소 세부 사항 관련하여 다른 의견 있는 학생들은 쌤에게 작성해서 보내주세요. ★★★★★

★★★★★ 평가 영역은 변경 불가하나, **평가 요소 세부 사항은 학생들의 의견을 반영하겠습니다!** ★★★★★

최고점 15점에서 위의 감점 요소를 반영한 후 반올림하여 성적에 반영합니다. 역시나 착하죠?

(ex. 감점 요소 1개인 경우 14.5 → 15점, 2개인 경우 14.0 → 14점,

3개인 경우 13.5 → 14점, 4개인 경우 13.0 → 13점)

<부록 6> 학생평가 포트폴리오 실제 자료: ⑩학생평가 전문성 개발활동
참여 기록 예시

2016학년도 (영어과) 전문적 학습공동체 운영 내용

1. 연수 개요

•연수 과정명	2016 학교안 전문적 학습공동체 직무연수		
•연수 주제	영어과 학생중심 배움중심수업 탐구		
•연수 장소	영어전용교실	•연수 기간	2016.9.5. ~ 2017.1.31.
•연수 총별	직무연수	•연수 시간	16시간
•대상	영어 교과 교사	•인원	7명

2. 운영내용

일정	시간	주제	세부내용
9월 7일 (수)	1	영어교과연계 진로교육 활성화 방안	- 1학년 : Holland test result를 활용하여 성격을 묘사하는 형용사를 익히고, 자신의 진로 탐색의 기회를 갖기 수업 방안 탐구 - 2학년 : 현재 또는 미래의 자신의 모습을 생각하고 영어로 자서전 쓰기 및 북아트 수업 방안 탐구
9월 21일 (수)	1	성격에 따른 직업세계 탐구수업을 위한 자료공유	- 1학년 : Holland 유형에 해당하는 직업들을 분류하고 영어자료 검색 및 발제, 자신의 미래직업 탐구를 위한 활동지 제작 - 2학년 : 자서전에 쓰이는 표현 및 내용을 모아 수업자료 제작
9월 28일 (수)	1	교육과정-수업-평가 일체화 방안 협의	- 교육과정 재구성 협의 - 공통 평가요소를 설정하고 그에 맞는 평가 문항 개발을 위한 협의
10월 4일 (수)	3	독서수업 활성화 방법 연수	- 선생님
10월 10일 (월)	1	학년별 배움중심수업 연계 평가 방안	- 9월 28일 협의 내용을 바탕으로 배움중심수업을 위한 활동지 제작 및 평가 기준 협의
10월 19일 (수)	1	영어교과 기초학력 책임교육 방안	- 학년별 기초학력 부족 학생의 학습과 관련된 정보 나눔 - 선발 된 학생들을 위한 구체적인 수업내용 구성과 학습자료 개발
10월 24일 (월)	1	학년별 독서수업 자료 공유 및 공동지도안 연구	- 1학년 : 퀴즈를 통한 독서수업자료 제작 - 2학년 : 독서 수행평가를 위한 활동수업 지도안 연구 및 활동지 제작 - 3학년 : 독서 수행평가를 위한 활동수업 지도안 연구 및 활동지 제작
11월 9일 (수)	1	인성연계 영어과 수업 방안 협의	- '고 월별 인성주제와 연계하여 수업시간에 활용해서 인성과 쉽고 누구나 시도해 볼 수 있는 영어수업 연계 노력으로 학생의 흥미와 관심 키우기 방안 협의
11월 30일 (수)	1	수업참관을 통한 수업나눔 및 공유	- 수업 짝 교사의 수업을 참관하여 인상 깊었던 점에 대해 이야기하고 배움

11월 30일 (수)	2	취약시기 인성연계 교과내 및 교과간 융합수업 방안 협의	- 1,2학년 : 영화 '도리를 찾아서' 를 활용하 여 자신의 단점을 생각해보고, 이를 극복할 수 있는 방법에 대해 설명하기 수업 방안 협 의
12월 14일 (수)	1	2017교육과정-수업-평가 일체화 방안	- 인성 및 독서교육을 바탕으로 교과재구성 방 안 모색 - 실생활에 필요한 영어 표현 및 내용을 중심 으로 학생 활동을 통한 영어 교육 방안 모색 - 수업시간 학생의 활동을 통한 평가기준 모색
12월 14일 (수)	2	혁신학교로서의 자체 종합평가를 위한 워크숍	- 혁신학교로서 '에서 진행 된 모든 교육과 정에 대한 평가 - 학생들의 역량 및 활동을 중심으로 혁신학교의 교육 내용에 대한 토론 및 2017 교육과정을 위 한 협의
계	16		

인성교육 연계교과 융합수업 방안 및 활동지

(영아 12월 30일 수업지도안)

구분	영아 12월 30일 수업지도안	영아 12월 30일 수업지도안
영아 12월 30일 수업지도안	영아 12월 30일 수업지도안	영아 12월 30일 수업지도안

인성교육 연계교과 융합수업 방안 및 활동지

(영아 12월 30일 수업지도안)

구분	영아 12월 30일 수업지도안	영아 12월 30일 수업지도안
영아 12월 30일 수업지도안	영아 12월 30일 수업지도안	영아 12월 30일 수업지도안

인성교육 연계교과 융합수업 방안 및 활동지

(영아 12월 30일 수업지도안)

구분	영아 12월 30일 수업지도안	영아 12월 30일 수업지도안
영아 12월 30일 수업지도안	영아 12월 30일 수업지도안	영아 12월 30일 수업지도안

인성교육 연계교과 융합수업 방안 및 활동지

(영아 12월 30일 수업지도안)

구분	영아 12월 30일 수업지도안	영아 12월 30일 수업지도안
영아 12월 30일 수업지도안	영아 12월 30일 수업지도안	영아 12월 30일 수업지도안

인성교육 연계교과 융합수업 방안 및 활동지

(영아 12월 30일 수업지도안)

구분	영아 12월 30일 수업지도안	영아 12월 30일 수업지도안
영아 12월 30일 수업지도안	영아 12월 30일 수업지도안	영아 12월 30일 수업지도안

인성교육 연계교과 융합수업 방안 및 활동지

(영아 12월 30일 수업지도안)

구분	영아 12월 30일 수업지도안	영아 12월 30일 수업지도안
영아 12월 30일 수업지도안	영아 12월 30일 수업지도안	영아 12월 30일 수업지도안

<부록 6> 학생평가 포트폴리오 실제 자료: ⑭교과 협의록 예시

(영어 독해와 작문)과 교 과 협 의 록

일자	2016년 11월 7일 월요일	장소	2층 본교무실																															
참석자	김OO, 김OO, 최OO, 홍OO																																	
협의 주제	2학기 2차지필평가 출제방안 및 출제범위																																	
협의 내용	<p>1. 출제 원칙</p> <p>가. 평가 문항의 출제는 타당도, 신뢰도, 난이도 등을 고려한다.</p> <p>나. 공동출제, 공동편집으로 학급 간 성적 차를 최소화한다.</p> <p>다. 전년도 기출문제, 참고서문제 등과 대조해 논란이 생기지 않도록 한다.</p> <p>라. 성적 불불리기에 해당하는 일은 하지 않도록 한다.</p> <p>마. 출제 관계 서류 관리 철저히 문제 사진 유출 예방을 철저히 한다.</p> <p>바. 단순한 암기테스트가 되지 않도록 종합적 사고를 요하는 문제를 출제한다.</p> <p>사. 시험시간 안에 풀 수 있는 문제로 난이도를 조절한다.</p> <p>아. 등급별로 고르게 분포할 수 있도록 변별도를 강화한다.</p> <p>자. 동점자를 최소화하도록 변별력 조정에 노력한다.</p> <p>2. 시험 시간 : 50분</p> <p>3. 교과목 평가 계획</p> <table border="1"> <tr> <th rowspan="2">교과</th> <th rowspan="2">학 년</th> <th rowspan="2">시험범위</th> <th rowspan="2">배점</th> <th rowspan="2">문항 수</th> <th colspan="2">서술형·논술형평가</th> <th rowspan="2">난이도</th> <th rowspan="2">예상 평균 점수</th> </tr> <tr> <th>반영 비율</th> <th>100점 만점 기준</th> </tr> <tr> <td>영어 독해와 작문A</td> <td>2</td> <td>-EBS올림포스 7~11, 13단원 -외부지문8개</td> <td>2.7점 ~4.5점</td> <td>15개</td> <td>7.7%</td> <td>22점</td> <td>상(20%) 중(60%) 하(20%)</td> <td>60/100</td> </tr> <tr> <td>영어 독해와 작문B</td> <td>2</td> <td>- 교과서 6단원, 8단원 -외부지문6개</td> <td>2.7점 ~4.5점</td> <td>15개</td> <td>7.7%</td> <td>22점</td> <td>상(20%) 중(60%) 하(20%)</td> <td>60/100</td> </tr> </table> <p>4. 답안지 사용 계획 : 일체형 답안지 사용(또는 “별지 답안지 사용” 이라고 기입)</p>					교과	학 년	시험범위	배점	문항 수	서술형·논술형평가		난이도	예상 평균 점수	반영 비율	100점 만점 기준	영어 독해와 작문A	2	-EBS올림포스 7~11, 13단원 -외부지문8개	2.7점 ~4.5점	15개	7.7%	22점	상(20%) 중(60%) 하(20%)	60/100	영어 독해와 작문B	2	- 교과서 6단원, 8단원 -외부지문6개	2.7점 ~4.5점	15개	7.7%	22점	상(20%) 중(60%) 하(20%)	60/100
	교과	학 년	시험범위	배점	문항 수						서술형·논술형평가				난이도	예상 평균 점수																		
반영 비율						100점 만점 기준																												
영어 독해와 작문A	2	-EBS올림포스 7~11, 13단원 -외부지문8개	2.7점 ~4.5점	15개	7.7%	22점	상(20%) 중(60%) 하(20%)	60/100																										
영어 독해와 작문B	2	- 교과서 6단원, 8단원 -외부지문6개	2.7점 ~4.5점	15개	7.7%	22점	상(20%) 중(60%) 하(20%)	60/100																										

(영어 독해와 작문)과 교 과 협 의 록

일자	2016년 11월 7일 월요일		장소	2층 본교무실																															
참석자	김OO, 김OO, 최OO, 홍OO																																		
협의 주제	2학기 2차지필평가 출제방안 및 출제범위																																		
협의 의의 내 용	<p>1. 출제 원칙</p> <p>가. 평가 문항의 출제는 타당도, 신뢰도, 난이도 등을 고려한다.</p> <p>나. 공동출제, 공동편집으로 학급 간 성적 차를 최소화한다.</p> <p>다. 전년도 기출문제, 참고서문제 등과 대조해 논란이 생기지 않도록 한다.</p> <p>라. 성적 부풀리기에 해당하는 일은 하지 않도록 한다.</p> <p>마. 출제 관계 서류 관리 철저히 문제 사전 유출 예방을 철저히 한다.</p> <p>바. 단순한 암기테스트가 되지 않도록 종합적 사고를 요하는 문제를 출제한다.</p> <p>사. 시험시간 안에 풀 수 있는 문제로 난이도를 조절한다.</p> <p>아. 등급별로 고르게 분포할 수 있도록 변별도를 강화한다.</p> <p>자. 동점자를 최소화하도록 변별력 조정에 노력한다.</p> <p>2. 시험 시간 : 50분</p> <p>3. 교과목 평가 계획</p> <table border="1"> <tr> <th rowspan="2">교과</th> <th rowspan="2">학 년</th> <th rowspan="2">시험범위</th> <th rowspan="2">배점</th> <th rowspan="2">문항 수</th> <th colspan="2">서술형·논술형평가</th> <th rowspan="2">난이도</th> <th rowspan="2">예상 평균 점수</th> </tr> <tr> <th>반영 비율</th> <th>100점 만점 기준</th> </tr> <tr> <td>영어 독해와 작문A</td> <td>2</td> <td>-EBS올림포스 7~11, 13단원 -외부지문8개</td> <td>2.7점 ~4.5점</td> <td>15개</td> <td>7.7%</td> <td>22점</td> <td>상(20%) 중(60%) 하(20%)</td> <td>60/100</td> </tr> <tr> <td>영어 독해와 작문B</td> <td>2</td> <td>- 교과서 6단원, 8단원 -외부지문6개</td> <td>2.7점 ~4.5점</td> <td>15개</td> <td>7.7%</td> <td>22점</td> <td>상(20%) 중(60%) 하(20%)</td> <td>60/100</td> </tr> </table>						교과	학 년	시험범위	배점	문항 수	서술형·논술형평가		난이도	예상 평균 점수	반영 비율	100점 만점 기준	영어 독해와 작문A	2	-EBS올림포스 7~11, 13단원 -외부지문8개	2.7점 ~4.5점	15개	7.7%	22점	상(20%) 중(60%) 하(20%)	60/100	영어 독해와 작문B	2	- 교과서 6단원, 8단원 -외부지문6개	2.7점 ~4.5점	15개	7.7%	22점	상(20%) 중(60%) 하(20%)	60/100
	교과	학 년	시험범위	배점	문항 수	서술형·논술형평가						난이도	예상 평균 점수																						
반영 비율						100점 만점 기준																													
영어 독해와 작문A	2	-EBS올림포스 7~11, 13단원 -외부지문8개	2.7점 ~4.5점	15개	7.7%	22점	상(20%) 중(60%) 하(20%)	60/100																											
영어 독해와 작문B	2	- 교과서 6단원, 8단원 -외부지문6개	2.7점 ~4.5점	15개	7.7%	22점	상(20%) 중(60%) 하(20%)	60/100																											
4. 답안지 사용 계획 : 일제형 답안지 사용(또는 “별지 답안지 사용” 이라고 기입)																																			

15. 자기평가보고서(Reflection Sheet)

한 학기동안의 학생평가 활동은 선생님께서 계획하신 대로 잘 이루어졌습니까? 잘 된 점 혹은 아쉬운 점 등을 구체적으로 적어주세요.

☞ 이번 학기는 모둠별 활동 수업을 목표로 성적을 고려해 각 반 모둠을 구성했습니다. 그리고 모둠 안에 학습 멘토를 두어 영어 과목에 있어서 멘토-멘티의 협동학습이 이루어지도록 노력했습니다. 또한 지필평가에 반영되는 EBS 올림포스 수능교재의 모둠별 수업 발표를 진행하였고 이에 대한 피드백을 전체 학생들이 작성해 각 모둠에 전달하면서 비공식적 평가가 수업 시간마다 진행되도록 했으며, 지필평가준비와 수업의 긴밀성을 높이고자 했습니다. 매 단원별 단어시험을 치루고 모둠별 평균 점수를 산출해 가장 준비를 잘해온 모둠과 그렇지 않은 모둠을 비공식적으로 기록했고, 이는 실질적인 점수에 반영되지는 않았으나 수행평가 중 한 항목인 단어수행평가(과제수행평가)를 효과적으로 준비시킬 수 있었고 학생들의 동기 부여 효과가 있었습니다. 쓰기수행평가 대비를 위해서 일주일에 한 번씩 영어원서읽기 시간을 활용해 학생들이 원서에 대한 리딩로그를 꾸준히 작성하게 하고 10회분의 리딩로그를 평가하고 쓰기수행평가를 치루었습니다. 전체적으로 지필평가를 비롯한 수행평가 준비를 평소 수업시간을 활용해 준비시키고 꾸준히 비공식적인 평가를 함으로써 수업과 평가 간의 긴밀성을 높이는 것이 목표였고, 이러한 점에서는 긍정적으로 평가합니다. 하지만 7개 반의 수업을 들어가면서 리딩로그와 같은 경우 매 시간마다 모든 학생에게 개별적 피드백을 제공하는 것에는 시간적으로도 많은 제약이 있었고 쓰기 수행평가의 피드백은 희망하는 학생들 위주로 제공할 수밖에 없었던 점이 아쉽습니다.

제출하신 자료를 바탕으로 할 때, 선생님의 학생평가 활동이 학생들의 학습 방법에 미치는 영향이 있다면 무엇입니까? 향후 교수활동 개선에 어떻게 활용하셨습니다?

☞ 학생 간 영어 실력차가 몹시 큰 공립고등학교라는 점을 고려해 볼 때, 영어 능숙도가 높은 학생들은 멘토활동이나 수업발표에 대해 굉장히 긍정적인 반응을 보였으나, 영어 능숙도가 낮은 학생들의 경우, 단원별 단어 시험에서 다른 모둠 친구들에게 해를 끼치지 않기 위하여 방새 단어를 외워오는 등 많은 노력을 하기도 했지만 많은 부담과 스트레스를 받았습니다. 수업 발표를 할 때에도 협동학습을 진행하고 역할을 분담하는 과정에서 모둠 구성원들 사이에 갈등과 협력이 이루어지기도 했습니다. 이번 학기 수업을 통해서 교사의 일방적 지식전달식 수업이 아니라 학생들이 자신의 학습 과정에 적극적으로 참여하도록 유도했고 수업 준비를 비롯해 수업 참여, 그리고 비공식적 평가까지 모두 주도해서 참여하도록 기회를 제공하였습니다. 이 방법은 영어 능숙도가 높은 학생들을 비롯해 영어 과목에 대한 학습 흥미도가 높은 학생들에게도 매우 긍정적으로 작용했습니다. 영어능숙도가 낮은 학생들의 경우도 멘토-멘티 활동에 적극적으로 참여한 학생들은 멘토 학생들로부터 배우며 덕을 보기도 했지만, 이러한 과정들은 비공식적으로 이루어진 활동들이기 때문에 교사의 적극적인 평가나 개입은 크지 않았습니다. 만일 7개 학급이 아니라 교사가 4개 정도의 학급만 담당하고 멘토-멘티 활동이나 쓰기 활동에 대한 개별적 피드백을 제공할 수 있게 된다면 더욱 효과적인 교수활동이 이루어지리라 사료됩니다.

선생님께서서는 지난 학기 공정한 학생 평가를 위해 어떠한 노력을 하셨습니까? 구체적으로 서술해 주세요.

☞ 영어능숙도가 높은 학생이든 낮은 학생이든 상관없이 학생들의 노력 여부와 수업 활동에 참여한 사항들을 비공식적으로 기록해 두었다가 생활기록부 교과 세부특기사항란에 기입을 해주었습니다. 지필평가와 같이 수치화 된 점수 결과로만 학생들을 평가하기 보다는 평소 수업 활동에 학생들이 참여할 수 있는 기회를 가능한 한 많이 제공하고자 노력했고, 이에 대한 비공식적 피드백과 기록, 교과 세부특기사항 입력과 같은 방식으로 인센티브를 제공하면서 학생들의 동기 유발을 유도하고 학생들의 노력에 대한 서술적 평가를 제공하고자 했습니다. 쓰기 수행평가와 같은 경우 채점 기준에 대해 동학년 영어 선생님들과 여러 차례의 회의를 거쳐 채점 기준을 명확히 하고 기준에 맞게 채점을 진행하였으며, 채점 후 바꾸어 재채점해보면서 채점자 간 신뢰도를 높이하고자 노력했습니다. 또한 채점 결과를 학생들에게 알려주고 개별적 피드백을 원하는 경우 채점 답안지를 보여주며 설명해 주는 방식으로 구두 피드백을 제공했습니다. 영어 원서 읽기 시간마다 진행한 리딩로그 작성 수행평가의 경우에는 첫 수업 시간에 명확한 기준과 방법을 설명해주는 시간을 가졌고, 학생들이 평가 방법 및 기준에 대해 명확하게 숙지하도록 노력했습니다.

-감사합니다-

[전문가용]

<부록 7> 교사의 학생평가 전문성 채점기준

I. 평가 계획

1. [교육과정]

※평가자료: ㉠ / ㉡ / ㉢ / ㉣ / ㉤

세부 기준 점수	평가 기준이 2009 개정 교육과정 에서 제시된 영어과 핵심성취기준 및 학습 요소를 포함함. ☞ 자료 ㉠ 참조	실제 수업 및 평가 자료가 교육과정(영 어과 핵심성취 기준)을 제대로 반영함. ☞ 자료 ㉡ / ㉢ / ㉣ / ㉤ 참조
1	모두 불만족(제대로 드러나지 않음)	
2~3	둘 중 한 항목만 만족 ※ 평가 기준이 영역별(듣기, 말하기, 읽기, 쓰기)로 영어과 핵심 성취기준을 구체적이고 명확하게 제시할 경우 3점	
4~5	두 항목 모두 만족 ※ 평가 기준이 영역별(듣기, 말하기, 읽기, 쓰기)로 영어과 핵심 성취기준을 구체적이고 명확하게 제시할 경우 5점	

2. [평가 목표]

※평가자료: ㉠ / ㉤ / ㉥

세부 기준 점수	학습 목표와 일치하는 평가 목표 를 제시함(평가 목표를 학습목표 와 관련하여 제시하고 있는가?) ☞ 자료 ㉠ / ㉤ 참조	평가 목표에 학습자들이 성취해야 할 도 달점 행동이 구체적이고 명확하게 드러 남. ☞ 자료 ㉠ / ㉥ 참조
1	모두 불만족(제대로 드러나지 않음)	
2~3	둘 중 한 항목 만족 ※ 비교적 구체적으로 제시할 경우 3점	
4~5	두 항목 모두 만족 ※ 비교적 구체적으로 제시할 경우 5점	

3. [평가의 실제성 (Practicality)]

※평가자료: ㉠ / ㉡ / ㉢ / ㉣

세부 기준 점수	교사의 평가에 대한 지식, 영어 교과에 대한 지식, 학습 진도, 업무 부담, 동료 교사와의 협력, 학교 여건 등을 고려할 때, 교사의 평가활동 준비 및 실행에 지나치게 많은 시간이 소요됨. ☞자료 ㉠ / ㉡ / ㉢ 참조	평가 과업 또는 평가 문항의 난이도가 학생 수준에 비해 지나치게 복잡하고 어려움. ☞자료 ㉡ / ㉢ / ㉣ '교과 협의록' 평가 결과 자료 참조	평가 과업 또는 평가 문항의 분량이 지나치게 길거나 짧음. ☞ 자료 ㉡ / ㉢ 참조
1	세 항목 모두 만족		
2	셋 중 두 항목 만족		
3~4	셋 중 한 항목만 만족 ※ 비교적 실제적인 평가 계획인 경우 4점		
5	세 항목 모두 불만족(해당사항 없음)		

4. [학습자의 인지적 특성에 관한 이해]

※평가자료: ㉠ / ㉤ / ㉨ / ㉣ / ㉥ / ㉦

세부 기준 점수	교사는 학습 준비도, 학습 스타일, 학습 전략 등 학습자들의 다양한 인지적 특성을 고려하여 개별적인 과제, 질문 및 형성평가를 제시하고 있음. ☞자료 ㉤ / ㉥ / ㉦ 참조	평가 계획서 및 교과 협의록의 내용을 바탕으로 할 때, 교사는 학생들의 영어 학습 방법 및 학습 활동에 영향을 미치는 요소들을 제대로 파악하고 있음. ☞자료 ㉠ / ㉨ / ㉣ 참조
1	모두 불만족(제대로 드러나지 않음)	
2~3	둘 중 한 항목만 만족 ※ 비교적 구체적으로 제시할 경우 3점	
4~5	두 항목 모두 만족 ※ 비교적 구체적으로 제시할 경우 5점	

5. [학습자의 정의적 특성에 관한 이해]

※평가자료: ㉠ / ㉡ / ㉢ / ㉣ / ㉤ / ㉥ / ㉦

세부 기준 점수	교사는 학습자들의 흥미와 학습동기 를 향상시킬 수 있는 평가 활동을 계획함 (예시>객관식 지필 평가 보다는 짝활동, 소그룹 활동 등의 다양한 활동 중심의 평가 활용). ☞자료 ㉠ / ㉡ / ㉢ / ㉤ 참조	교사는 학급 내 특별한 학습 요구를 지닌 학생을 파악하고 그 특성을 고려하여 평가 활동을 계획 하고 있음(예시>시험 불안, 주의력 결핍 증세가 있는 학생들 또는 장애학생들에 대한 고려). ☞자료 ㉠ / ㉣ / ㉤ / ㉦ 참조
1	모두 불만족(제대로 드러나지 않음)	
2~3	둘 중 한 항목만 만족 ※ 비교적 구체적으로 제시할 경우 3점	
4~5	두 항목 모두 만족 ※ 비교적 구체적으로 제시할 경우 5점	

II.평가 도구 개발

6. [평가 도구 이해]

※평가자료: ㉠ / ㉡ / ㉢ / ㉤ / ㉥ / ㉦ / ㉧ / ㉨

세부 기준 점수	평가계획서,수행평가·지필평가(선택형·서술형)의 문항 및 채점기준을 고려 할 때, 평가 목표에 적합한 평가 도구 유형을 선택함 . ☞자료 ㉠ 평가계획서 참조	평가 도구의 장·단점을 제대로 이해 하고 있음. ⁴⁾ ☞자료 ㉡ / ㉢ / ㉤ / ㉥ / ㉦ / ㉧ / ㉨ (1)참조
1	모두 불만족(제대로 드러나지 않음)	
2~3	둘 중 한 항목만 만족 ※ 비교적 구체적으로 제시할 경우 3점	
4~5	두 항목 모두 만족 ※ 비교적 구체적으로 제시할 경우 5점	

4) ※평가 도구의 장·단점 이해의 예시

- 고등사고 능력 평가를 위해 객관식, 단답식 평가 보다는 서술형 평가 방식을 활용함.
- 협동적 문제 해결 능력 평가를 위해 그룹 과제 또는 평가 과업을 제시함.
- 논술형 문항의 경우 답안 예측이 어려워 평가 실시 전 채점기준표 또는 부분 점수 기준표를 평가 실시 전에 미리 작성하기 어렵거나 채점자간 신뢰도에 문제가 발생할 가능성이 있음.

7. [평가 도구 개발]

※평가자료: ㉠ / ㉡ / ㉢

세부 기준 점수	영어과 핵심성취기준의 달성여부를 점검할 수 있는 평가 과제 및 문항을 개발하여 제시함. ☞자료 ㉠ / ㉡ / ㉢ 참조	언어 영역별(듣기, 말하기, 읽기, 쓰기) 특징을 고려하여 평가 과제 및 문항을 개발하여 제시함. ☞자료 ㉠ / ㉢ 참조
1	모두 불만족(제대로 드러나지 않음)	
2~3	둘 중 한 항목만 만족 ※ 비교적 구체적으로 제시할 경우 3점	
4~5	두 항목 모두 만족 ※ 비교적 구체적으로 제시할 경우 5점	

8. [평가 도구 선택]

※평가자료: ㉠ / ㉡ / ㉢ / ㉣ / ㉤

세부 기준 점수	표준화된 평가 도구를 그대로 활용함(예시> EBS 듣기 평가). ☞자료 ㉠ / ㉡ / ㉢ / ㉣ 참조	표준화된 평가도구를 평가 목적에 따라 변형하여 사용함(예시> 교과서 문제, 학습지, 수능 기출문제, 모의고사 문제 활용 등). ☞자료 ㉠ / ㉡ / ㉢ / ㉣ / ㉤ 참조
1	모두 불만족(제대로 드러나지 않음)	
2~3	둘 중 한 항목만 만족 ※ 비교적 구체적으로 제시할 경우 3점	
4~5	두 항목 모두 만족 ※ 비교적 구체적으로 제시할 경우 5점	

9. [평가 도구 관리]

※평가자료: ㉠ / ㉡

세부 기준 점수	<p>지난 평가의 문항별 특성(문항 유형별 변별도, 난이도 등)과 평가도구의 양호도(신뢰도, 타당도) 개념을 고려하여 해당 평가 도구 개발에 반영함.</p> <p>☞자료 ㉠ / ㉡ 문항 출제 관련 협의록 자료 참조</p>	<p>해당 평가 결과 분석을 바탕으로 차후 평가 도구 개발에서 고려할 사항을 구체적으로 제시함.</p> <p>☞ 자료 ㉡ 참조</p>
1	모두 불만족(제대로 드러나지 않음)	
2~3	<p>둘 중 한 항목만 만족</p> <p>※ 평가 도구의 특성(신뢰도, 타당도 포함 문항 양호도 등)을 구체적으로 검토할 경우 3점</p>	
4~5	<p>두 항목 모두 만족</p> <p>※ 평가 도구의 특성(신뢰도, 타당도 포함 문항 양호도 등)을 구체적으로 검토할 경우 5점</p>	

10. [채점기준]

※평가자료: ㉠ / ㉡

세부 기준 점수	평가 실시 전 수행평가 및 지필형(서술형 평가 포함)평가의 채점기준을 미리 설정하여 제시하였음.	설정된 채점기준은 비교적 구체적이고 명확하게 제시되는가?	채점기준에 인정 답안을 포함하여 제시하고 있는가?
1	모두 불만족		
	※ 채점기준 미작성		
2	셋 중 한 항목만 만족		
3~4	<p>셋 중 두 항목 만족</p> <p>※ 비교적 구체적으로 나타나는 경우 4점</p>		
5	세 항목 모두 만족		

Ⅲ. 평가 실행

11. [평가 시기]

※평가자료: ㉠ / ㉡ / ㉢ / ㉤ / ㉥ / ㉦ / ㉧

세부 기준 점수	평가 계획에 따라 적절한 시기에 실제 평가활동이 이루어짐. ☞ 자료 ㉠ ‘평가 시기’ 와 ㉡, ㉢, ㉤ ‘실행 시기’ 비교	평가의 교육적 목적을 극대화시키기 위 해 적절한 시기에 평가가 실시됨(수업 활동 중간에 질문, 퀴즈, 형성 평가 등 다양한 평가가 적절하게 실시되는가?) ☞ 자료 ㉥ / ㉧ 참조
1	모두 불만족(자료 미제출 포함)	
2~3	둘 중 한 항목만 만족 ※ 비교적 구체적으로 나타나는 경우 3점	
4~5	두 항목 모두 만족 ※ 비교적 구체적으로 나타나는 경우 5점	

12. [학생 안내]

※평가자료: ㉤

세부 기준 점수	①평가 계획(시기 및 시간), ②평가 도구(평가 과업, 과제 등), ③채점기준, ④ 평가 실시 방법, ⑤범위 등을 학생들에게 안내함.
1	미리 안내하지 않음.
2	1개 항목만 안내함.
3	2개 항목만 안내함.
4	1~2개 항목을 제외하고 안내함.
5	모두, 구체적으로 안내함.

13. [채점 및 성적 부여]

※평가자료: ㉠ / ㉡ / ㉢ / ㉣ / ㉤ / ㉥ / ㉦ / ㉧

세부 기준 점수	채점 방식이 학습 목표 및 평가 목표 에 적합함. ☞자료 ㉠ / ㉢ / ㉤ 자료 참조, ㉡ 채점기준과 ㉤, ㉦의 실제 채 점된 답안 비교	문항별 채점 또는 교사 간 교차 채점을 통해 정확한 채점을 실시하고자 함. ☞자료 ㉠ / ㉣ / ㉤ 교과 협의록에서 채 점 관련 내용/ ㉤ / ㉦ 참조
1	모두 불만족(자료 미제출 포함)	
2~3	둘 중 한 항목만 만족 ※ 비교적 구체적으로 나타나는 경우 3점	
4~5	두 항목 모두 만족 ※ 비교적 구체적으로 나타나는 경우 5점	

14. [숙제 및 형성평가 실시]

※평가자료: ㉠ / ㉢ / ㉤ / ㉭

세부 기준 점수	교사는 학생들의 학습 목표 도달 여부를 판단할 수 있도록 해당 수 업 내용과 연계되는 숙제 또는 형 성평가를 제시함. ☞자료 ㉠ / ㉢ / ㉤ 참조	교사는 다양한 비형식적 평가 방법을 이 해하고 실시함(예시>학습태도 체크리스 트, 과제 제출 체크리스트, 관찰, 면담, 학습 일지, 오답노트 등). ☞자료 ㉢ / ㉭번 참조
1	모두 불만족(자료 미제출 포함)	
2~3	둘 중 한 항목만 만족 ※ 비교적 자주 제시할 경우 3점	
4~5	두 항목 모두 만족 ※ 비교적 자주 제시할 경우 5점	

15. [평가 참여]

※평가자료: ㉞ / ㉟ / ㊱

세부 기준 점수	학생들에게 자기 평가의 기회를 제공함	학생들에게 동료 평가의 기회를 제공함
1	모두 불만족	
2~3	둘 중 한 항목만 만족 ※ 비교적 여러 번 기회를 제시할 경우 3점	
4~5	두 항목 모두 만족 ※ 비교적 여러 번 기회를 제시할 경우 5점	

IV.평가 결과 분석 및 활용

16. [결과 분석 및 해석]

※평가자료: ㊱ / ㊲

세부 기준 점수	평가 문항에 관한 측정오차, 평균, 표준편차 등 평가 결과에 관한 기 술통계 분석을 실시함.	양적 분석 결과를 바탕으로 질적 분석 결과(평가에 대한 해석)를 제시함.
1	모두 불만족 ※교과 협의록 미제출	
2~3	둘 중 한 항목만 만족 ※ 비교적 구체적으로 제시를 제시할 경우 3점	
4~5	두 항목 모두 만족 ※ 비교적 구체적으로 제시를 제시할 경우 5점	

17. [평가 결과 활용]

※평가자료: ㉔ / ㉕ / ㉖ / ㉗ / ㉘

세부 기준 점수	진단평가, 형성평가 및 총 괄평가의 분석 결과를 통해 전체 학생들의 학업 성취도 에 대한 종합적인 판단을 내림. ☞자료 ㉔ / ㉕ / ㉖ 참조	평가 결과를 바탕으로 개별 학생들의 학습에 대해 일관 성 있고 종합적인 판단을 내림 (예시>성적통지표에서 의 교사의 코멘트나 학생생 활기록부의 교과별 세부영 역 기입 내용을 볼 때, 학 생들의 성취도에 대해 지속 적이고 종합적인 기록을 유 지함). ☞ 자료 ㉗ / ㉘ 참조	언어 영역별로 평가 결과에 서 나타난 문제점을 진단하 고 원인을 분석한 후 차기 평가 활동의 개선방안을 제 시함. ☞자료 ㉕ 참조
1	모두 불만족 ※ 교과 협의록 미제출		
2	셋 중 한 항목만 만족		
3~4	셋 중 두 항목 만족 ※ 비교적 구체적으로 나타나는 경우 4점		
5	세 항목 모두 만족		

18. [학부모와의 의사소통]

※평가자료: ㉙ / ㉚

세부 기준 점수	평가 실시 전 학부모에게 구체적인 평 가 목적 및 평가 계획에 관한 정보를 제공함. ☞자료 ㉙ 가정통신문 배부 또는 학교 홈페이지 게시 내용 또는 학 급 안내지 참조	평가 실시 후 학생의 언어 기능별 학습 곤란 도, 과거 평가 결과와의 비교 및 분석, 영어 학습상의 문제점 진단 등 평가 결과에 관한 정보를 제공함(예>평가 활동, 평가 결과 및 피드백에 관한 가정 통신문을 배부하거나 홈 페이지 또는 블로그 활용, 오답노트 활용 등) ☞ 자료 ㉚ 참조
1	전혀 드러나지 않음 ※자료 미제출 포함	
2~3	둘 중 한 항목만 만족 ※비교적 구체적으로 제시할 경우 3점	
4~5	두 항목 모두 만족 ※비교적 구체적으로 제시할 경우 5점	

19. [학생 피드백]

※평가자료: ㉔ / ㉕ / ㉖ / ㉗ / ㉘ / ㉙

세부 기준 점수	평가 결과에 대한 양적, 질적 피드백을 제공함.	평가 결과에 관해 학생들에게 개별적인 피드백을 구체적으로 제공함.
	☞자료 ㉔ / ㉕ / ㉖ / ㉗ / ㉘ / ㉙ 참조	☞ 자료 ㉙번 참조
1	제공하지 않음 ※자료 미제출 포함	제공하지 않음 ※자료 미제출 포함
2	양적인 피드백만 제시하거나 일부 ※ 3점의 두 항목 중 하나만 만족	학생들에게만 개별적인 피드백을 제공함
3	양적인 피드백(예시> 체크리스트에 표시, 도장, 채점, 기호 사용 등)만 제공할 뿐 질적 피드백은 제공하지 않음.	일부 학생들에게만 개별적인 피드백을 제공함
4	대체적으로 양적, 질적 피드백을 함께 제공하거나 비교적 모든 학생들에게 개별적인 피드백을 제공함 ※ 5점의 두 항목 중 하나만 만족	
5	대부분의 평가 결과에 대한 양적 피드백과 질적 피드백(학습에서의 어려움 진단 및 개선점에 관한 개별적인 코멘트)을 함께 제시함.	비교적 모든 학생들에게 개별적인 피드백을 제공함

20. [교수 기술 및 전략]

※평가자료: ㉚ / ㉛-(2)

세부 기준 점수	평가 결과를 교수 기술 및 전략에 반영함.
1	평가 전과 이후의 수업에 변화가 없음. ※교수 기술 및 전략에 관한 언급이 전혀 없음
2~3	평가 결과를 바탕으로 교사 자신의 교수활동 전반에 대해 평가하며, 개선이 필요한 부분을 파악하나 구체적으로 활용방안을 제시하지는 않음.
4~5	평가 결과를 차후 수업 내용, 수준, 수업 자료, 교수 방법 등의 조절에 적극적으로 활용하는 방안을 구체적으로 제시함.

V. 평가에 대한 태도 및 인식

21. [평가의 공정성]

※평가자료: ㉠ / ㉡ / ㉢ / ㉣

세부 기준 점수	학생들의 성적에 직·간접적으로 영향을 줄 수 있는 교사의 행위를 최소화하였음(예시>공정한 평가를 위해 평소의 학생평가 활동을 꼼꼼하게 기록, 활용함). ☞자료 ㉠ 참조	개별 학생 특성을 고려하여 모든 학생들에게 공정한 평가 기회를 부여하였음(예시> 장애 학생들에 대한 배려). ☞자료 ㉠ / ㉡ 참조	지난학기 공정한 학생평가 활동을 위한 교사의 노력이 지속적이고 구체적으로 나타남. ☞자료 ㉢-③ 참조
1	모두 불만족 ※ 자료 미제출		
2	셋 중 한 항목만 만족		
3~4	셋 중 두 항목 만족 ※ 비교적 구체적으로 나타나는 경우 4점		
5	세 항목 모두 만족		

22. [평가 활동에 대한 반성]

※평가자료: ㉠ / ㉣

세부 기준 점수	한 학기동안 스스로 자신의 평가 활동을 반성하고 문제점을 발견함. ☞ 자료 ㉠ / ㉣-① 참조	발견된 문제점에 대해 스스로 개선 방안을 제시함. ☞자료 ㉠ / ㉣-② 참조
1	모두 불만족	
2~3	둘 중 한 항목만 만족 ※평가활동에 관한 교사의 자기 평가 내용을 비교적 구체적이고 상세하게 제시할 경우 3점	
4~5	두 항목 모두 만족 ※평가활동에 관한 교사의 자기 평가 내용을 비교적 구체적이고 상세하게 제시할 경우 5점	

23. [전문적 자율성]

※평가자료: ㉞

세부 기준 점수	평가 관련 연수, 워크숍, 대학강의, 컨설팅, 교사 커뮤니티, 세미나, 스터디그룹 등의 활동에 참여 횟수
1	전혀 없음(미제출)
2	학기당 1회
3	학기당 2회
4	학기당 3회
5	학기당 4회 이상 참여

24. [동료교사와의 협업능력]

※평가자료: ㉠ / ㉞ / ㉡

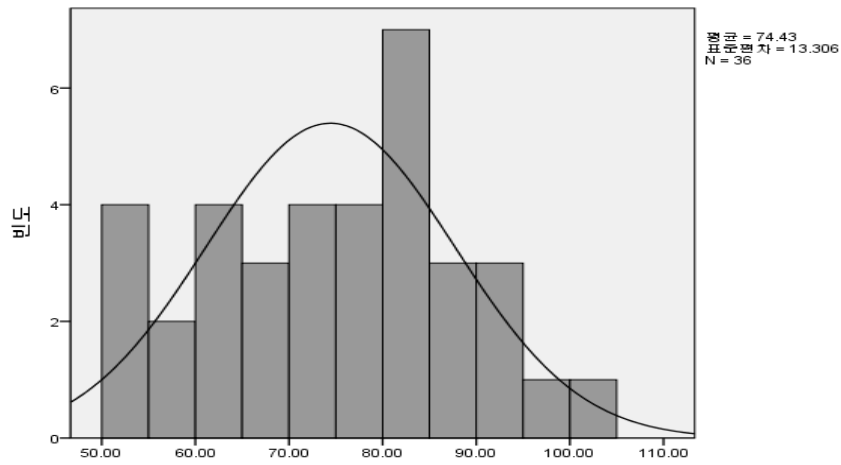
세부 기준 점수	[평가실시 전] 서술형 평가, 수행평가의 유사 답안 및 부분점수 부여에 관한 기준 설정을 위한 동료 교사 간 협의를 하는가?	[평가실시 중] 채점기준이나 부분점수 기준에 대한 동료 교사와의 이견이 있을 경우 원인을 분석하고 해결방안을 제시하는가?	[평가실시 후] 동료 영어교사들과의 교과 협의회 또는 학년 협의회나 학교 밖 교사모임(전문적 공동체)을 통해 평가 결과에서 발견된 문제점을 분석하고 해결방안(교수·학습 방법의 개선)을 공유하는가?
1	모두 불만족 ※교과협의록 미제출 포함		
2	셋 중 한 항목만 만족		
3~4	셋 중 두 항목 만족 ※비교적 구체적으로 제시할 경우 4점		
5	세 항목 모두 만족		

25. [평가의 환류적 기능]

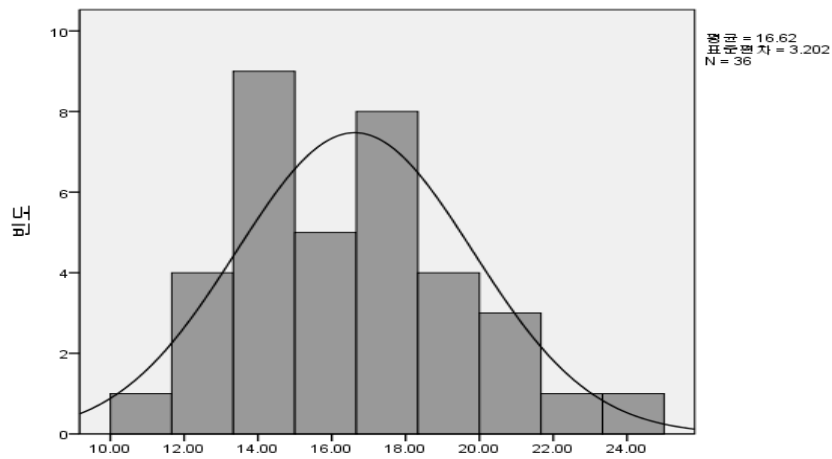
※평가자료: 전체 자료

세부 기준 점수	교사의 전반적인 학생평가 활동이 학습자들의 학습과 관련된 인지 및 정서적 영역에 긍정적인 영향을 주는 방향으로 계획, 반성적으로 실천됨.
1	평가 활동의 주요 목적이 다른 학생과의 학업 성취도의 비교에만 있는 경우
2~3	교사의 전반적인 평가 활동이 학습자들의 인지적 또는 정의적 영역에 부분적인 변화를 가져오는 경우(예시>교사의 평가활동의 일부가 학생들의 의사소통 능력 향상에 도움이 되었음). ※비교적 구체적, 명시적으로 드러날 경우 3점
4~5	교사의 평가 활동이 학습자들의 인지적(학습 방법, 전략 등), 정의적 영역(학습 흥미, 태도 등)에 모두 긍정적인 변화를 가져오는 경우(예시>교사가 평가 결과에서 발견된 개별 학생들의 영어 학습의 강점과 약점을 바탕으로 학습 개선을 위해 구체적인 피드백을 제공하거나, 교수학습활동을 개선하기 위해 지속적으로 교사 전문성을 계발하고 동료들과 협력함). ※비교적 구체적, 명시적으로 드러날 경우 5점

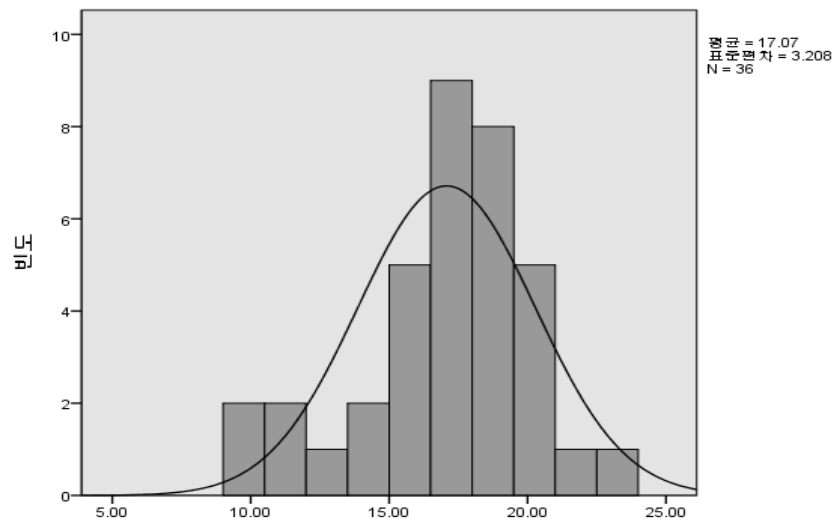
<부록 8> 교사의 학생평가 전문성 점수 분포(히스토그램)



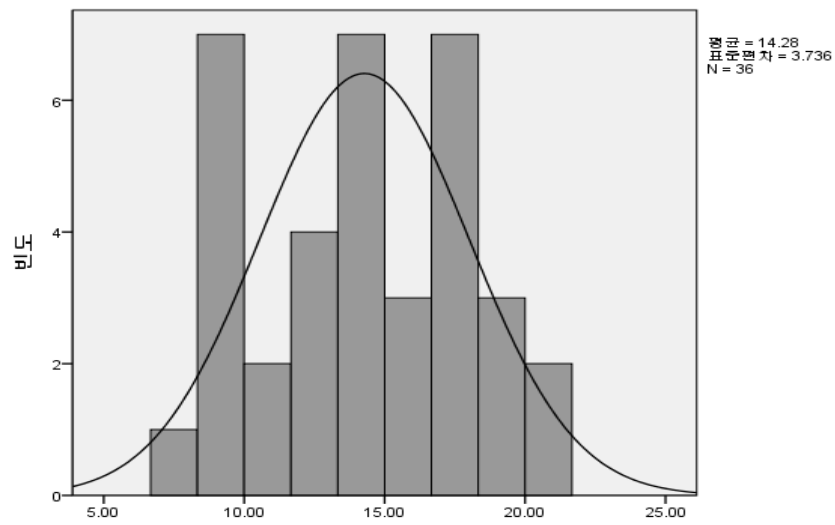
[그림 IV-1] 교사의 학생평가 전문성 총점



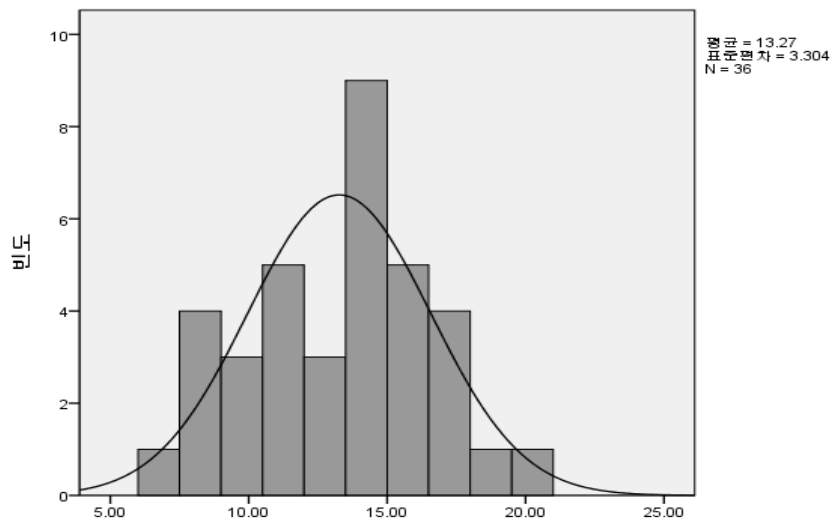
[그림 IV-2] '평가계획' 영역의 점수분포



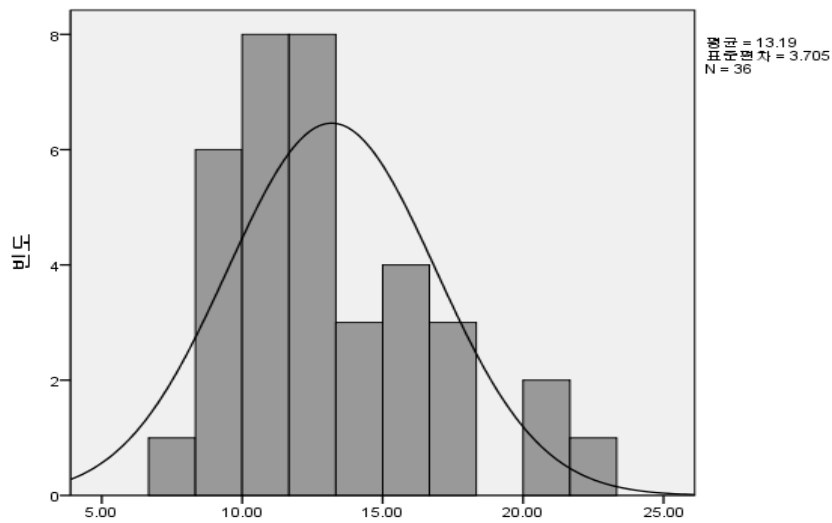
[그림 IV-3] ‘평가도구 개발’ 영역의 점수분포



[그림 IV-4] ‘평가 실행’ 영역의 점수분포



[그림 IV-5] ‘평가 결과 분석 및 활용’ 영역의 점수분포



[그림 IV-6] ‘평가에 대한 태도 및 인식’ 영역의 점수분포

The Development and Validation of an Assessment Instrument for Secondary English Teachers' Competencies in Student Assessment

A Dissertation for the Degree
of
Doctor of Philosophy in Education
by

Nayoung Kim

Major Advisor: Hyun-Jeong Park (Ph.D.)

Department of Education
The Graduate School
Seoul National University

2017

Abstract

The Development and Validation of an Assessment Instrument for Secondary English Teachers' Competencies in Student Assessment

Nayoung Kim
Department of Education
The Graduate School
Seoul National University

Recently there has been a shift in public attention towards a variety of education reforms to respond to changes in future society brought on by the fourth industrial revolution. There have also been many changes in educational evaluation worldwide. As the paradigm of educational evaluation changes from 'Assessment of Learning' to 'Assessment for Learning' and 'Assessment as Learning', a new role and awareness change is required for teachers, who play a key role implementing these changes in schools. Since the full-scale introduction of free semester, achievement standards-based assessment and the 2015 National curriculum in Korea, teachers' competencies in student assessment have been continuously emphasized in order

to understand the meaning of student assessment as an integrated activity involving teaching, learning and assessment, and to carry out assessment activities in a more valid and reliable manner. In particular, as the 2015 National curriculum focuses on students' participation in the teaching and learning process and process-focused assessment, academic interest in teachers' competencies in student assessment has increased. In addition, teachers' practical capacity in assessing students has been more highly emphasized. In this context, it is worthwhile to develop a valid and reliable assessment instrument to evaluate the level of teachers' competencies in assessing students objectively in school practices and support the improvement of teachers' competencies in student assessment on the basis of the results.

This study aims to develop such an assessment tool for evaluating secondary English teachers' authentic competencies in student assessment and implement a validation study of the tool. For this purpose, the assessment scale, components of student assessment portfolios and a scoring rubric for expert evaluation were developed based on the analysis of relevant literature and expert discussion. First 'student assessment portfolios' that reflect teachers' assessment techniques were collected in a semester long evaluation of 36 middle school and high school English teachers working in Seoul, Gyeonggi and

Incheon. Then three assessment experts evaluated the assessment portfolios using the assessment scale through rater training. After expert evaluation, a written survey about the properness of the assessment instrument was implemented by English teachers and assessment experts. Finally, based on the evaluation results, the validity, reliability and properness of the developed assessment tool were tested.

The main results in the study are summarized as follows.

Firstly, the assessment scale for secondary English teachers' competencies in student assessment consist of five sub-constructs: 'assessment planning', 'development of assessment tools', 'assessment implementation', 'analysis and utilization of assessment results', and 'attitudes and consciousness on assessment'. Furthermore, each construct includes five specific evaluation standards respectively. When comparing the average scores of each component, the scores for 'development of assessment tools' were the highest, while those of 'analysis and utilization of assessment results' and 'attitudes and consciousness on assessment' were relatively lower than other components. When it comes to specific evaluation standards, the average scores of 'assessment practicality' were the highest, while those of 'assessment participation' and 'communication with parents' were the lowest.

Secondly, the content validity of the developed assessment tool was analyzed by five experts in English education. It was shown that the 25 specific standards appropriately measured five sub-constructs of teachers' competencies in student assessment.

After analysis it was determined that the developed assessment tool provided an acceptable level of reliability. When the internal reliability of the evaluated items were analyzed using classical test theory (CTT), each component was shown to have a Cronbach's alpha coefficient of .56~.78, with a total coefficient of .88. Correlation coefficients of .84~.98 among the results of evaluation by three experts also showed a high level of inter-rater consistency. Furthermore, an acceptable level of reliability was generally obtained using generalizability theory. The Generalizability coefficient was .8743 and the Dependence coefficient was .7961, representing above .8 and .7 respectively.

Finally, according to the analysis of a written survey on the properness of the assessment tool, evaluation using assessment portfolios provided teachers with opportunities for self-reflection about their own student assessment activities, which presumably can be more practical and effective alternatives to teacher training related to student assessment for the improvement of teachers' competencies in student assessment. It should be noted that difficulties in data collection for assessment portfolios and

the multi-faceted evaluation of teachers' competencies in student assessment based only on the collected assessment portfolios were limitations of the assessment tool. The results imply that the developed assessment instrument can be utilized more effectively in the teacher training curriculum in order to enhance teachers' authentic competencies in student assessment if some of these problems are addressed.

To put it briefly, because only 36 teachers' assessment portfolios were analyzed in this study, it was difficult to generalize the results and some limitations were discovered in the properness of the assessment tool. Nonetheless, the assessment tool using teachers' assessment portfolios needs to be used more extensively in order to improve teachers' authentic competencies in student assessment as an alternative approach to overcome the limitations of prior teacher education related to assessment.

*** keywords:** teacher's competency in student assessment,
authentic assessment activities, assessment
portfolios, development of assessment tool,
validation study

*** *Student Number:*** 2013-31074